Digitized by the Internet Archive
in 2015

https://archive.org/details/creativityimpuls00ward

Department of Psychology

CREATIVITY AND IMPULSIVITY
IN KINDERGARTEN CHILDREN

by

William C. Ward

Department of Psychology
Duke University

DATE: April 7, 1966

Approved:

Michael A. Wallach

Michael A. Wallach, Supervisor

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Psychology in the Graduate
School of Arts and Sciences of Duke University

ABSTRACT

(Psychology-Developmental)

CREATIVITY AND IMPULSIVITY
IN KINDERGARTEN CHILDREN

by

William C. Ward

Department of Psychology
Duke University

DATE: April 7, 1966

Approved:

Michael A. Wallach

Michael A. Wallach, Supervisor

Harold Schiffman

Jack J. Preiss

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department
of Psychology in the Graduate School of
Arts and Sciences of Duke University

ABSTRACT

Conceptualization of creativity as a trait present in normal individuals raises the question of how early in childhood it can be identified. Substantial intercorrelation among measures of divergent thinking, and relative independence of measures of convergent thinking, are sufficient to identify the trait; while study of intercorrelations with other psychological domains and relations to creativity studied in adulthood are necessary for assessing its usefulness. The present study concentrates on whether a creativity trait can be found in kindergarten children, and investigates its relationship to a cognitive style dimension, reflection--impulsivity. Hypothesized relationships are based on an information-processing interpretation of variables proposed to underly individual differences in creativity and cognitive impulsivity.

Ss were 87 kindergarten children of diverse IQ and SES. They were tested individually, one procedure to a session, in a game-playing rather than test-taking atmosphere. Three creativity procedures were administered: suggesting possible interpretations of simple abstract figures; naming instances of a category; and naming uses for common objects. Forms A and B of the Peabody Picture Vocabulary Test (PPVT) provided the convergent intelligence measure. Five measures of impulsivity were

obtained. Two, given as ability tests, were similar to Kagan's figure-matching tasks; while the PPVT and an additional procedure gave three impulsivity measures in a nonevaluative context.

Three sets of specific hypotheses were tested. For creativity, it was hypothesized that the divergent thinking procedures would show significant intercorrelations, for both fluency and uniqueness measures; and that these measures would not show significant correlation with mental age (MA) on the PPVT. Predictions for impulsivity were that latencies on all five tests would show significant intercorrelation and independence of MA; while errors would be negatively related to both latencies and intelligence. Finally, two predictions across domains were made: *Ss* high on creativity and low on intelligence should be highly impulsive (have short latencies) when impulsivity is measured under evaluative conditions; and *Ss* high on intelligence and low on creativity should be highly reflective (long latencies) on all impulsivity measures.

Creativity measures based on the Uses and Instances Naming procedures intercorrelated significantly and were independent of MA in both sexes; but the figure-interpretation measures were unrelated to these and, in males, showed some dependence on MA. Excluding *Ss* with IQ below 100 increased creativity intercorrelations for males, while failing to change relationships for females. Subsidiary analyses suggested that common processes, notably degree of involvement, may underly performance on all creativity procedures; but response totals from only two

of these conform to hypotheses. Special problems associated with the figure-interpretation procedure might account for its failures to correlate, despite the presence of a unitary trait.

Hypotheses concerning impulsivity received detailed confirmation. Latencies were significantly intercorrelated and independent of MA; errors were negatively related to both latencies and MA. Impulsivity shows generality across a range of cognitive tasks and is not systematically affected by administration in a game-like as opposed to an ability-testing context.

Only tentative assessment of relationships across domains could be made, since little unity appeared in the creativity results. Hypotheses were not confirmed, but several suggestive relationships were found: Males high on the figure-interpretation procedure and low on intelligence were highly impulsive on all measures; for both sexes and all creativity measures, *S*s high on both creativity and intelligence formed the most reflective subgroup.

Possible interpretations of results were considered. It is clear that the creativity measures are something other than unusual tests of convergent intelligence; it remains to be shown that they measure a unitary trait or that their scores depend on the same processes that account for creative performance in older individuals.

ACKNOWLEDGEMENTS

It is startling to complete that most independent of projects, the dissertation, and then begin to count the number of people who have contributed to its execution. My greatest debt is to Dr. Michael Wallach, my advisor and chairman - for many ideas and several measures; for prepublication access to his manuscript; and for appropriately alternating doses of support and criticism. Drs. Jerome Kagan and the late Albert Amons gave highly relevant stimulation; Dr. Kagan also contributed a measure to the study. My dissertation committee members, Drs. Borstelmann, Kimble, Preiss, and Schiffman, were helpful in various critic and sounding-board roles.

Teachers, parents, and children gave exceptional cooperation. I am particularly grateful to Mrs. Barbee, Mrs. Boineau, Miss Hinman, Mrs. Hutchins, and Mrs. Jones, directors of the kindergartens at which I worked and pretested; all were most helpful and amazingly tolerant of great disruption.

Several individuals contributed to the analysis of data. I especially want to thank Helen Brantley, for her analysis of questionnaire data; and my wife Mary, for much of the analysis of creativity data.

Members of the secretarial staff at Duke and at Stanford

saved me from the typing of various manuscript drafts. The National Science Foundation saved me from starvation, via two years of fellowships. Research money was provided by one of these fellowships, and the Duke University Computer Center supplied free computer time.

Finally, Mary Ward helped as wife and as typist; somehow she managed to surmount the extra demands a dissertation puts on both these roles.

W.C.W.

CONTENTS

ABSTRACT iii

ACKNOWLEDGEMENTS vi

LIST OF TABLES ix

I. INTRODUCTION 2

Studies of Creativity, 4
Some Efforts at Interpretation, 44
Reflection--Impulsivity, 53
Creativity and Impulsivity, 63
Age Differences, 70
Aims and Hypotheses of the Present Investigation, 75

II. PROCEDURE 79

Subjects, 79
Testing Conditions, 82
Procedures, 84
Questionnaire, 91
Scoring of Creativity Instruments, 92

III. RESULTS 95

Creativity, 95
Reflection--Impulsivity, 129
Creativity and Impulsivity, 135

IV. DISCUSSION AND CONCLUSIONS 143

Reflection--Impulsivity, 143
Interpretation of Creativity Results, 148
Conclusions, 158

REFERENCES 165

LIST OF TABLES

I. Guilford's Divergent Thinking Factors 13

II. Five Experiments Employing Guilford-like Creativity Procedures 16

III. Subject Characteristics 81

IV. Categorization of Responses in Creativity Tasks 97

V. Median Item-Sum Reliability Coefficients 98

VI. Mean Creativity Scores 100

VII. Creativity and Intelligence Correlations for all Subjects 102

VIII. Creativity and Intelligence Correlations for Males 104

IX. Creativity and Intelligence Correlations for Females 106

X. Intelligence Means and Standard Deviations 110

XI. Creativity and Intelligence Correlations for Males Excluding those with IQ less than 100 111

XII. Creativity and Intelligence Correlations for Females Excluding those with IQ less than 100 112

XIII. Correlations among Time in Task for Creativity Procedures 117

XIV. Correlations between Time and Response Measures 118

XV. Correlations between Responses Accepted and Those Discarded 123

XVI. Category Scores on Uses - Correlations with Fluency and Mental Age Measures 126

XVII. Reflection--Impulsivity Intercorrelations and Correlations with Intelligence for Males 131

XVIII. Reflection--Impulsivity Intercorrelations and Correlations with Intelligence for Females 132

XIX. Reflection--Impulsivity for Male Creativity--Intelligence Subgroups Based on Pictures Performance 138

XX. Reflection--Impulsivity for Male Subgroups Based on Pictures Performance - Summary of Analyses of Variance 139

XXI. Reflection--Impulsivity for Male Creativity--Intelligence Subgroups Based on Instances and Uses Performance 141

CREATIVITY AND IMPULSIVITY
IN KINDERGARTEN CHILDREN

CHAPTER I

INTRODUCTION

The research reported here grew out of an interest in two recently discussed dimensions of individual difference--"creativity," as measured by Wallach and Kogan (1965); and "reflection--impulsivity," as measured by Kagan and his collaborators (Kagan, Rosman, Day, Albert, and Phillips, 1964). Each dimension has been studied with middle class elementary school children; each can be clearly distinguished from conventional IQ measures in the samples studied; and each claims to involve important aspects of the child's mental functioning.

The major purpose of the present investigation was to provide a strong test of the replicability of these two dimensions. Children of kindergarten age, ranging from upper middle to working class in background, served as subjects. If consistent individual differences in "creativity" and "impulsivity" could be found despite these differences from the conditions under which the dimensions were first identified, good support would be given for their pervasiveness and generality. In particular, it would be clear that they were not limited in applicability to children of upper middle class background, and that they were

not solely a function of intellectual socialization by the school environment.

A second purpose was to investigate interrelationships of creativity and impulsivity. An interpretation of Wallach and Kogan's results led to the expectation of congruences between the two sets of work. This part of the study, of course, was contingent on successful replication of both of the dimensions in question.

There is little relevant background literature for Kagan's work. Creativity, on the other hand, has been studied in many guises and from a variety of perspectives. Therefore, a major part of the introductory discussion will be devoted to an analysis of research on creativity. This discussion is biased toward emphasis on that work which has clearest implications for the study of creativity in children; toward raising issues rather than adding one more summary to the literature; and toward setting the research by Wallach and Kogan in proper perspective to other attempts at objective measurement of creativity.

Following this discussion, Wallach and Kogan's results, and to some extent those obtained by Mednick and his coworkers (1962), are restated in the language of "cognitive controls." The reflection--impulsivity dimension is then discussed, and its possible relationships to the creativity findings are considered. Next, reasons for particular interest in investigating these relationships with preschool children are suggested. Finally, the present experimental investigation is described.

Studies of Creativity

General Discussion. The creativity literature is characterized by a great diversity of approaches. Several recent reviewers (Golann, 1961; Mooney, 1963) have categorized these in terms of interest in the products of creative activity; in the personality of the creative individual; in the conditions suitable for creation; and in the creative process. In the first category would come the study by C. Taylor (1959), in which fifty-six measures were obtained on a group of research scientists; including, for example, number of publications, professional memberships, peer and supervisor ratings. Twenty-seven factors emerged in an analysis of these scores, indicating the "almost overwhelming complexity of the criterion problem" (Golann, 1963, p. 549)--and pointing, perhaps, to the futility of such a "shotgun" approach.

Studies of the personality of the creative individual offer greater promise of psychological meaningfulness. In writings by Freud and neo-analytic theoreticians, the creativity at issue has implicitly been that of the productive artist or writer. Freud (1947, 1959a) saw the forces leading to creativity as being the same as those which in some individuals lead to neurosis; thus, he is in part responsible for the long life of the notion that genius and madness are closely related. Kris emphasized the interplay of preconscious processes with those of the unconscious (1952); while, in Kubie's Neurotic Distortion of the Creative Process (1958), preconscious processes are of

predominant importance, and creativity is seen in conjunction with psychological health rather than sickness. Neo-analytic writings have thus come closer to those of other personality theorists who would discuss creativity in a context of "self-actualization" (Rogers, 1954; Goldstein, 1939). Closely related to the latter is Schactel's (1959) discussion of creativity as involving perceptual openness to the environment.

These discussions suggest some interesting possibilities, but generally remain at the level of speculation. Still, a few studies have been conducted, attempting to show that individuals rated as "creative" have more primary process thought available, along with more effective control over this form of thinking, than have less creative persons (e.g., Myden, 1959; Holt and Havel, 1960; Pine and Holt, 1960). The major question arising from these formulations is whether "creativity"--artistic productivity--is associated with psychological health or with its opposite.

Also within the realm of personality approaches to creativity is work by MacKinnon, Barron, and others at the Institute of Personality Assessment and Research. MacKinnon (1962), for example, has studied creative architects, research scientists, and mathematicians, choosing individuals who are nominated as creative by fellow practitioners of their fields. A variety of techniques have been used to find attributes which distinguish these individuals from less successful ones in the same areas. It should be clear that "creativity" as employed here refers to no single or small set of abilities. MacKinnon notes that

"The successful and effective architect must, with the skill of a juggler, combine, reconcile, and exercise the diverse skills of businessman, lawyer, artist, engineer, and advertising man, as well as those of author and journalist, psychiatrist, educator, and psychologist." (1962, p. 486) Selection is thus for individuals who are outstanding in many respects.

Other writers have concentrated on the conditions of creativity, usually in terms which relate the process of creating to effective problem solving. Typical emphases include reference to the necessity for a playful attitude toward the subject matter and to a need for freedom from externally imposed pressures (Ghiselin, 1952; Hadamard, 1954). The four-stage analysis of problem solving formulated by Wallas in 1926 provides the central theme of many analyses. The necessity for preparation, the period of incubation, the moment of illumination, and the work of verification--these are seen to apply to creativity as to ordinary problem solving, if indeed a distinction is maintained. Explanations center about what goes on to precede the creative moment, the moment of illumination; they range from unconscious integration (Ghiselin, 1952); to decay of interfering associations set up by excessive concentration (Woodworth and Schlosberg, 1954). Discussions of this sort frequently seem designed to embrace creativity without extension of the problem-solving model advanced by the discussants (e.g., Newell, Shaw, and Simon, 1958; Asher, 1963).

While these various types of study are interesting in their own right, there is some question as to the usefulness of

applying the term "creativity" to many of them. If "creativity" is synonymous with psychological health, effectiveness in problem solving, or the production of many scientific papers, the label is of little value; it can be replaced by less pretentious terminology without loss of communication. Of recent work, MacKinnon's seems least susceptible to this criticism: his subjects are indeed "creative," in popular interpretations of that term; and there is no indication that a few less honorific variables could summarize their characteristics.

None of this work provides much direction for research on creativity in children. With the exception of the occasional prodigy, children do not create the finished, high-level products required. There are, however, several programs of research in which creativity is studied at a more basic level, that is, at the level of a dispositional or ability trait. I. Taylor (1959) has characterized definitions at this level as concerned with "expressive creativity"--where skills and product quality are relatively unimportant, and emphasis is on the expressive freedom of the individual. Here, then, there is the attempt to measure a characteristic which may be necessary, though not sufficient, for adult productivity of the sort that is popularly labeled creative.

This approach has obvious utility, in that creativity of this lowest level may not be limited to adults, nor to a few particularly outstanding individuals; rather, it may provide meaningful differentiation over a broad range of age and ability. Further, it is a reasonable guess that greater unity of

psychological process can be found with such an approach than when higher-level definitions are employed; with the latter, superior intelligence, high motivation, and sufficient opportunity to acquire and to utilize skills and knowledge will be among the requisites for creative production.

One difficulty that must be faced by studies at the trait level is that of the lack of an external validating criterion. This is particularly a problem when discussion centers on the creativity of children. No self-evident tie exists between "expressive creativity" in the five or ten year old and outstanding artistic or scientific achievement in the adult. A reasonable, though quite difficult, course for investigation would be an attempt to demonstrate that such a tie can be found: that the trait which can be measured in the child or in the sophomore is a necessary asset if he is ever to enjoy exceptional productivity. A less ambitious aim, however, seems legitimate. This is to demonstrate that a dimension of individual difference, at the level of "expressive creativity," can be measured reliably; that this dimension can be distinguished from other dimensions, such as that measured by the IQ test; and that it has broad enough implications for behavior to justify consideration as a distinct aspect of the cognitive structure. If relationships to creativity in the sense of outstanding productivity can be found, well and good; but such ties are not necessary to make such a dimension a meaningful one.

In these requirements, the criterion of face validity is the easiest to meet. Reasonable candidates include ideational

fluency (Wallach and Kogan, 1965); sensitivity to gaps and missing elements (Torrance, 1964); freedom from self-imposed restriction on the form of problem solution (Guilford, 1956); and many others. Support for any of these choices can be obtained by examination of the introspections of acknowledged productive thinkers (e.g., Ghiselin, 1952). More difficult is the demonstration that the abilities or dispositions measured do in fact have sufficient coherence to justify giving them a common label. Thorndike has set the problem as follows:

> "When we use any term to designate an attribute of an individual, be it abstract intelligence, sociability, or creativity, the distinct label implies that there exists a set of behaviors of the individual that can be grouped together because they exhibit a common quality. Furthermore, if the label is a useful one, the set of behaviors that it designates can be differentiated from the sets designated by other labels ...We may appropriately ask how well the attribute 'creativity' meets these joint criteria of designating a reasonably extensive set of behaviors which (1) have some degree of coherence and (2) can be distinguished from other sets of behaviors." (Thorndike, 1963, pp. 44-45)

Other writers agree to the appropriateness of these requirements. Wallach and Kogan (1965), for example, asked the question. "...could one demonstrate the existence of greater and lesser degrees of a cognitive capability that was like intelligence in regard to being a pervasive, broad dimension, but yet was independent of intelligence, and which could appropriately be labeled 'creativity'?" (p. 287)

Critics adopting these criteria have shown that, in many investigations, the so-called creativity measures lack substantial intercorrelation; and in fact these measure correlate with one another no more strongly than with general intelligence

indicators.[1] Where this is so, there is no warrant for claiming that something distinct from IQ is being measured. It does not follow that a dimension appropriately labeled creativity cannot be demonstrated; several recent investigations have had considerable success in doing so. Before discussing these, however, it will be useful to review the two major programs of research which have provided the starting point for much of the work currently under attack.

Guilford. The most influential work has been that of Guilford and his collaborators. These investigators have used factor-analytic techniques in a number of studies of the "aptitudes of high-level personnel." Over the course of a decade, a theoretical model of intellect has evolved, encompassing three dimensions on which intellectual abilities are categorized. Each cell defined by the intersection of values on these dimensions is filled by a factor which has been either demonstrated or postulated. As of the 1959 version of the model, 47 of 90 postulated factors have actually appeared. (Guilford, 1959a)

The three dimensions are concerned with kind of process or operation involved, type of content, and kind of product required. The dimension of chief interest involves five operations: cognition, memory, convergent thinking, divergent thinking, and evaluation. Guilford has commented:

---

[1] It should be emphasized that in discussing relationships between creativity and intelligence, the latter term is used with the relatively restricted meaning of "those abilities which standard IQ tests measure."

> "In view of the active nature of creative performances, the production aspects or steps are most conspicuous and probably most crucial. Among the productive-thinking abilities another logical distinction appears. With some productive-thinking factors and the tests that measure them, thinking must at some time converge toward one right answer; the significant type of thinking involved has been called 'convergent thinking.' With other productive-thinking factors and their tests, thinking need not come out with a unique answer; in fact, going off in different directions contributes to a better score in such tests. This type of thinking and these factors come under the heading of 'divergent' thinking. It is in divergent thinking that we find the most obvious indications of creativity." (Guilford, 1957, pp. 111-112)

Divergent thinking factors, however, are not the only ones important for creativity:

> "Creative thinking;...depends upon different combinations of factors, and the combination of factors significant to the task will vary from time to time. The problem confronting us here, as with problem solving, is to recognize the main categories of creative production and to seek the signficant combinations of factors involved in them." (Guilford, 1956, p. 289)

The factors suggested to be of particular importance to creative thinking are listed in Table 1, along with several examples of tests loading principally on each factor. Four factors involve some kind of fluency in producing words, ideas, or complex expressions; three (including Redefinition) involve flexibility in conceptual activity; and one involves novelty, uniqueness, or originality of productions. Guilford summarizes as follows:

> "Considering the factors in the divergent-thinking category together, it is obvious that the freedom to change direction of thinking varies considerably from one instance to another. Different degrees of situation-imposed restriction are involved. But generally, within whatever limits that are imposed by external restrictions, the need for rejecting

> or superseding a response and for trying or producing a new one is the common element in this group of factors. There is also a difference in the amount of self-imposed restriction or freedom. This depends upon the individual rather than upon the situation. It is largely in this source of variation that we find the divergent thinking factors." (1956, p. 281)

Guilford's convergent thinking tests coincide roughly with those on which conventional IQ measures are based. Despite the factorial differences which can be found in this domain, various IQ tests have much of their variance in common; full-scale IQ's, rather than subparts of the domain, have been the more useful for predictive purposes. (McNemar, 1964) Guilford's comments on his measures of divergent thinking, quoted above, suggest clearly that these tests should have a common element, one which is distinct from that shared by tests of convergent thinking. It seems a fair test, then, to apply Thorndike's criteria in asking, first, to what extent Guilford's creativity tests are intercorrelated; and second, whether this intercorrelation is greater than that between the convergent and divergent thinking measures.

Thorndike (1963) subjected the results of the study by Wilson, Guilford, Christensen, and Lewis (1954) to a reanalysis. Each of the eight divergent thinking and six convergent thinking factors was represented by the two tests loading highest on it. The average correlations between factors within and between the convergent and divergent domains were assessed. Within the convergent thinking domain, the average r was .23; within divergent thinking, .14; and between domains, .12. The divergent

Table I

Guilford's Divergent Thinking Factors[1]

Factor	Explanation and Examples of Tests Loading on Factor
Word Fluency	Ability to produce words fulfilling structural requirements, with no emphasis on meaning: anagrams; producing words with specified first letter.
Associational Fluency	Production of words to meet specific requirements of meaning: find a word which is a synonym for each of a pair of words; as, for both prevaricate and recline.
Ideational Fluency	Production of ideas to fulfill certain requirements in limited time: naming objects that are hard, white, and edible; naming uses for bricks; writing appropriate titles for a short story plot. Scored for fluency, not quality, of responses.
Expressional Fluency	Rapid juxtaposition of words to meet requirements of sentence structure; ability to put ideas into words, rather than to have ideas; completing similes with different words; writing sentences for the frame: "W_c_e_n_."
Spontaneous Flexibility	Ability or disposition to produce a great variety of ideas, with freedom from inertia or perseveration, where test does not require jumping from one category to another: brick uses scored for number of category shifts.
Adaptive Flexibility	Flexibility in problems where this is necessary for solution: matchsticks problems.
Redefinition	Ability to give up old interpretations of familiar objects in order to use them or their parts in some new way: "Which of the following could best be adapted to making a needle: pencil, radish, fish?"

---

[1] Drawn from Wilson, Guilford, Christensen, and Lewis, 1954; Guilford, 1956, 1959a, 1959b.

Table 1 (cont.)

Factor	Explanation and Examples of Tests Loading on Factor
Originality	Giving responses that are unusual: bricks uses; clever responses: titling story plots; and/or remote responses: giving the consequences to follow from some event, with scoring for "non-obvious" consequences.

thinking tests, then, correlated no more highly with one another than with standard convergent tests. On the basis of the correlations obtained, Thorndike suggested that Word Fluency and Adaptive Flexibility might belong to the convergent rather than the divergent domain. With this shift, average correlations become .24 within convergent thinking, .19 within divergent thinking, and .08 between domains. Such a pattern of correlations would constitute weak evidence for distinguishability of two domains; but there is no basis in Guilford's theorizing for transferring just these two factors to the convergent domain.

My own analysis of another set of data (Guilford, Frick, Christensen, and Merrifield, 1957), using a procedure identical to Thorndike's, showed a similar lack of coherence for the divergent thinking domain. Correlations across the three factors within convergent thinking averaged .24; those within divergent thinking averaged .18; and those between domains averaged .15.

Finally, there are similar patterns in studies conducted by other investigators but using modifications of Guilford's tests. Results for five studies, employing as subjects students from the sixth grade through the last year of high school, are shown in Table 2. Ss in three studies were highly superior to population norms, while in the other two they were of about average intelligence. In no case were the correlations within "creativity" tests substantially higher than those between "creativity" and "intelligence" tests, and in two cases they were actually lower.

Table II

Five Experiments Employing Guilford-like Creativity Procedures

Study	Subject Characteristics			Results	
	Age	Sex	Mean IQ	Mean Correlation Within Creativity	Mean Correlation Between Domains
Piers et al., 1960	Grades 7-8	M&F	127	0.26	0.19
Cline et al., 1962	High School	M&F	100	0.22	0.34
Getzels & Jackson, 1962	Grades 6-12	M&F	144	0.30	0.27
Cline et al., 1963	High School	M&F	100	0.23	0.34
Flescher, 1963	Grade 6	M&F	138	0.11	0.04

Averaging results from the seven studies discussed, mean correlation among creativity measures is .20; mean correlation between creativity and IQ-like measures is .21. There is no indication, then, that Guilford's divergent thinking factors define a broad, unitary domain, or that they have any variance in common that is not shared with IQ.

The possibility remains that some set of these tests will define a domain which is relatively independent of conventional intelligence indexes. Three sets of factors may be suggested as candidates. One would involve the fluency factors; another, the flexibility factors; and a third, the lone originality factor. Each of these has some face validity as an index of "creativity." Some evidence from Guilford's batteries and from other studies can be raised to examine the intercorrelations of each of the three.

Three flexibility factors appeared in each of two analyses by Guilford. For each factor in flexibility and in convergent thinking, the two tests loading highest were chosen as representative. Correlations calculated for the domain included those within as well as between factors. In one study (Wilson et al., 1954), correlations within flexibility averaged .13, while those between flexibility and convergent thinking averaged .15. In the other (Guilford, Frick, Christensen, and Merrifield, 1957), within correlations averaged .24; between, .24. In each case, correlations within the convergent domain were on the order of .30. No unity appears, then, for the flexibility domain.

The construct of originality, for which there is a single

factor in Guilford's system, shows more coherence than does flexibility, but is strongly related to conventional intelligence indexes. In three different factor analyses, the two tests best loading on originality have average correlations of .25, .19, and .01 with convergent thinking (Guilford, Kettner, and Christensen, 1956; Guilford et al., 1957; Wilson et al., 1954). In a study by Piers, Daniels, and Quackenbush (1960), four tests of originality chosen from Guilford's batteries all intercorrelated significantly, with a median correlation of .38. Each of these, however, was significantly correlated with the IQ measure employed, with a median value of .32. Finally, in a study by Barron (1956) a number of indexes of originality, including three of Guilford's, showed significant intercorrelations, and several predictions as to differences between relatively more and relatively less original individuals were confirmed. Staff ratings of originality correlated highly with a composite test index. Unfortunately, no indexes of IQ were employed.

It appears, then, that the flexibility dimension shows no internal consistency; and that originality shows fairly high correlations with IQ. For fluency, the evidence is somewhat ambiguous. Guilford's correlations were examined, again choosing the two best tests to represent each factor, and taking all correlations within the domain. In one analysis, in which four fluency factors appeared, mean correlation within fluency was .29, while that between fluency and convergent tests was .24 (Guilford et al., 1956). In another, with three fluency factors,

the within correlations averaged .23, while those between averaged .14 (Wilson et al., 1954). In a third, where only ideational fluency appeared, between correlations averaged .01 (Guilford et al., 1957).

If Thorndike's suggestion that Word Fluency be removed from the divergent domain is taken, then, in the first study above, the mean correlation within creativity becomes .31; that between domains, .22; in the second study, the figures are .28 and .08. The evidence is weak, but suggests that fluency factors other than Word Fluency may have a degree of intercorrelation that is not due to variance in common with conventional IQ.

Examination of a number of factor analyses of Word Fluency carried out by other investigators, most of them preceding the work by Guilford, indicates that this factor is quite clearly correlated with a prototypical convergent ability, Verbal Comprehension. Those studies using oblique rotations have reported substantial correlations among these factors (Thurstone and Thurstone, 1941; Carroll, 1941). In one battery Thurstone finds a correlation of .51 between Word Fluency and Verbal Comprehension; in a more refined battery, a correlation of .42. In those studies which report the correlation matrix, in every case the correlations among fluency tests are of the same order of magnitude as those between Fluency and Verbal Comprehension. (Thurstone and Thurstone, 1941; Carroll, 1941; C. Taylor, 1947) Such evidence as is available from other studies also points to substantial correlations between Verbal Comprehension and Word Fluency. (Johnson and Reynolds, 1941; Cattell, 1946)

Two studies which report correlation matrices isolate a fluency factor other than Word Fluency. When the correlations within this factor are compared with those between its tests and tests of verbal comprehension, there is again no clear separation. In Carroll's study (1941), four tests highly loaded on a factor described as a combination of Ideational and Expressional Fluency had average intercorrelations of .19, while correlating with Verbal Comprehension with a mean value of .21. In C. Taylor's study (1947) a similar factor had correlations within averaging .43; those between, .25.

Gewirtz (1948a, b) investigated verbal fluency in children aged 5-0 to 6-7, using his own set of seven procedures. These bear more similarity to Word Fluency than to those for the others of Guilford's factors. Mean intercorrelation of fluency tests was .35; while mean correlation between fluency and Stanford-Binet MA was .46.

Gewirtz also had two observers rate the children on 15 selected scales from the Fels Child Behavior Scales. Four scales were discarded, since the raters failed to show significant agreement on these at the .05 level. The remaining ratings were quite variable, with .55 the median reliability coefficient. Despite this unreliability, MA showed significant positive correlations with five of the scales: originality, lack of social apprehensiveness, competitiveness, curiosity, and gregariousness. Median correlations of the rating scales with fluency scores were significant for these five scales, and for no others. Hence, no possibility for separation of fluency from general intelligence

appears here.

Summarizing the results of the studies examined here, there is little indication that a dimension of individual difference distinct from convergent intelligence has been isolated. Almost all the evidence involves correlations within creativity and between creativity and intelligence; only in Gewirtz's study were other variables investigated in conjunction with these. In no case was a clear separation of divergent from convergent intelligence obtained. This was true particularly when the divergent thinking domain was treated as unitary.

When subparts of divergent thinking were examined, flexibility factors were found to hold little variance in common. Tests of originality intercorrelated more highly, but were just as strongly correlated with IQ as with one another. Within the fluency domain, Word Fluency appeared to be indistinguishable from Verbal Comprehension. For other fluency factors, a possible separation from convergent intelligence could not be ruled out entirely; but neither was there any strong basis for claiming its existence.

Before the discussion turns from Guilford's work, several potential contributors to these negative findings deserve consideration. One of these concerns the range of talent in the samples tested. It has been suggested that a triangular relationship may obtain between creativity and intelligence. McNemar (1964) proposed that there will be found a very low incidence of creativity in subjects of average and below average intelligence, while for subjects above average there will be

a wide range of creative talent. Other investigators have suggested a cutoff point at around an IQ of 120 such that, below this level, measures of creativity and those of IQ will be positively related; while, above this level, they will be relatively independent. (McKinnon, 1960; Roe, 1960)

Guilford's subjects, in the main, have been Air Force and Navy enlisted personnel and officer candidates. No mean IQ's are reported for these groups, but it is likely that they fall within the average range. It could be, then, that a more select group is required to achieve independence. However, it should be noted that lack of intercorrelation of creativity measures is as serious a problem in Guilford's results as is their correlation with measures of convergent intelligence; demonstrating independence from the latter would not be sufficient without showing also that the divergent thinking domain becomes more unitary when low IQ subjects are excluded. The available studies which do report IQ levels fail to support this proposition. Those by Flescher (1963), Piers _et al_. (1960), and Getzels and Jackson (1962), all performed with children of superior intelligence, produced results similar to those of studies with average IQ subjects. A study by Guilford, Merrifield and Cox (1961), performed with ninth graders, also indicated positive creativity-IQ relationships across a wide range of ability level. A group covering an unselected range of IQ showed lower mean performance on twelve of fifteen creativity tests than did subjects with IQ ranging from 120 to 130; while the latter group had lower scores on ten of the fifteen tests than had a group with IQ greater than 130.

Wallach and Kogan (1965) have suggested that several features of the testing conditions employed in creativity assessment may have contributed to positive relationships with IQ. First, creativity has been measured in a context of ability testing; Guilford's tests, in fact, have been given under military discipline. (Guilford, 1956) Such conditions may inhibit freedom and spontaneity of thought. The subject may be led to search for answers that are certain to be right, by criteria of logic and conventionality, rather than for answers which may be good though unconventional. If the kind of thinking to be assessed is that which "goes off in different directions," rather than converging "toward one right answer," the measurement context should be one in which the similarity to ability testing situations is minimized. The study of Wallach and Kogan, to be discussed below, provides a clear indication of how this aim can be effected.

One feature of the testing context is important enough to merit mention at this point. This is that of the short time limits imposed in Guilford's measures--typically, two minutes or less per item. Guilford has indicated on numerous occasions the importance he assigns to speed of production; for example, in Guilford and Christensen (1956) he comments: "Extreme care was used in timing the tests, since most of the test scores are largely a function of speed." (p. 20) Elsewhere, he defines Ideational Fluency as "the rapid generation of units of verbal or semantic information;" Associational Fluency is "the rapid generation of semantic units to fulfill a relationship." (Guilford, 1963, p. 33)

Great concern with response speed has been a highly prominent feature of attempts to measure convergent intelligence--to an undue extent, according to some critics. For example:

> "I.A. Taylor (1959) argues that the 'intelligence' is an invention of western culture and selects and stresses the values important in our society. He maintains that intelligence tests 'essentially concern themselves with how fast relatively unimportant problems can be solved without making errors.' (p. 54) 'In another culture,' he continues, 'intelligence might be measured more in terms of how adequately important problems can be solved, making all the errors necessary and without regard for time." (Torrance, 1962, pp. 18-19)

If there can be question as to the appropriateness of speed as criterial to intelligence assessment, certainly there is greater question as to its validity in assessment of creativity. Wallach and Kogan (1965) have argued that power, rather than speed, should be criterial. Clearly this is the case for higher-level productivity, where an outstanding achievement may require years for preparation and execution. For lower-level production, empirical support is available for expecting later rather than first products to represent the creative capacity of the individual. In one study, Bousfield and Barkley (1950) found uncommonness of restricted verbal associates to increase with time since beginning of the task. Christensen, Guilford, and Wilson (1957) report a similar result for a test of Ideational Fluency. Finally, Mednick's (1962) associative hierarchy notions, discussed below, provide theoretical grounds for the expectation that those who produce many responses in a short time period may even be less productive over longer intervals than subjects whom they outstrip early in the task.

A final point worth noting is that, in some cases, the failure of creativity tests to intercorrelate may be simply a function of the great diversity of tests employed. McNemar (1964) is perhaps too extreme when he claims that "...so little is known about the creative process that measuring instruments are, seemingly, chosen on a trial-and-error basis" (p. 876); nonetheless, there are examples in which measures long regarded as belonging to other domains have been appropriated for the study of creativity. Getzels and Jackson (1962), for example, used an embedded figures test as part of their creativity battery. It was shown above, however, that subparts of Guilford's divergent thinking domain do not show appreciably greater unity than does the domain as a whole; it is unlikely, then, that test diversity is a sufficient explanation for the low correlations within the domain.

Torrance. Turning from Guilford's work and that related to it, another large program of research on creativity has been conducted by Torrance and his associates. Early procedures used by the group were modifications of Guilford's tests, such as substituting tin can uses for bricks uses, to make them more appropriate for use with young children. The tests more recently used have moved in the direction of open-endedness, requiring more complex products which are scored by multiple criteria to give a number of indexes. The list includes, among others, indexes of fluency, flexibility, cleverness, "inventiv-level," constructiveness, and adequacy. These have generally been summed without regard to their intercorrelations to give

a composite creativity score.

Both verbal and nonverbal tests have been used. A typical nonverbal test requires the examinee to fill in detail to complete designs on a page of circles or other figures. Verbal tests include Impossibilities--name as many impossible things as you can think of; and Product Improvement--name ways in which a nurse kit, for example, could be improved to make it more fun for children to play with.

Relatively complete descriptions of the measures employed are given in Torrance (1962). It is clear from the examples cited that more than divergent thinking is required; abilities which Guilford assigns to the convergent thinking or evaluation domains must figure prominently in performance. Torrance has recognized explicitly that diverse kinds of behavior are being assessed, and that it may not be legitimate to try to represent all of the abilities by any single index (1962, p. 17) Nonetheless, this is the procedure that has frequently been followed. Nowhere has he published the correlations among measures upon which this index is based, though it is stated that "There are low correlations between verbal and nonverbal creative abilities and they appear largely independent." (Torrance and Gowan, 1963, p. 3; cited in Wallach and Kogan, 1965, p. 7)[1]

Torrance has shown great concern with the question of

---

[1] Sherman (1965) used measures based on those of Torrance with nine year old children. No correlation matrix was reported; however, factor analysis led to pure test factors showing orthogonality to one another.

whether creativity can be distinguished from intelligence. He has in several cases taken the top twenty percent of a group in terms of the summed creativity measure, taken a similar twenty percent in terms of intelligence quotients, and examined the overlap between these two groups. The typical report is that the top twenty percent on intelligence tests includes only thirty percent of the top group on creativity tests, a finding which would indicate that the domains are relatively independent (Torrance, 1962). Without information on creativity intercorrelations, however, this finding is of little import.

The major effort to demonstrate the distinctiveness of creativity from intelligence has been exerted in studies which attempted to show that individuals scoring high on creativity tests but not high on those of intelligence perform as well on tests of academic achievement as do individuals high on intelligence and low on creativity. This result was claimed in six of eight samples of grade school and college students. (Torrance, 1960) It appears, however, that the creative but "low intelligent" subjects who perform as well as their highly intelligent peers are actually of above average intelligence. Two of the successful studies were performed with graduate student groups. In the others, the mean IQ's for low intelligent groups were 127, 121, 112, and 122 (Torrance, 1962, Table 1, p. 58) There is no demonstration, then, that creativity is contributing to performance; rather, two highly intelligent groups are shown to achieve similarly. Further, in the two elementary schools in which "highly intelligent" children showed greater achievement

than "highly creative" ones, mean IQ for the latter was substantially lower than in the groups cited above--98 in one case, and 107 in the other.

More recently, Torrance has amended his argument. There is postulated to be a critical level of IQ above which conventional intellectual abilities do not contribute substantially to academic achievement. The suggested cutoff point is around an IQ of 120 (Yamamoto, 1961), in agreement with suggestions by Roe (1960) and MacKinnon (1960). MacKinnon's suggestion, however, was that above an IQ of 120, conventionally assessed IQ makes little or no contribution to creative performance, while other variables account for individual differences in attainment. Torrance argues, rather, that above an IQ of 120, IQ will make no difference, but creativity status will have an effect, *on academic achievement* (Yamamoto, 1961). Other writers would question such a proposal. For example, Flescher writes, in discussing a study with Guilford-like tests in which creativity scores showed no correlations with achievement:

> "A compelling question regarding research on creativity is why original thinking should be expected to be a relevant factor in scholastic achievement, the nature of the latter being so markedly convergent in content. Formal learning is very unlike original and creative activity, and convincing evidence of a significant relationship is still wanting. At present, intelligence tests are referred to as tests of *scholastic aptitude*. On the basis of current knowledge, it is an unwarranged assumption to broaden the concept of scholastic aptitude to include creative potential." (Flescher, 1963, pp. 264-265)

The evidence to support the critical level proposal is inadequate and frequently irrelevant. One set of six studies

performed by Yamamoto (1961) chose, in each case, the top twenty percent on creativity of the school group studied. Each group was divided into three parts, on the basis of intelligence scores. In three cases out of the six, the expected results were obtained: academic achievement was greater for subjects with IQ above 120 than for those with IQ less than 120, while achievement did not differ significantly for groups in which mean IQ was greater than 120. Creativity scores likewise were superior for those over 120, without significant differentiation between groups over 120 in IQ. Some support, then, is given for the proposition that creativity and intelligence are independent above an IQ of 120, while not so below this level. An alternative, of course, is the possibility that one or both of these assessment devices is not validly differentiating subjects who are far above the norms in ability. At any rate, this experiment seems to offer no evidence directly pertinent to the question of a positive relation between creativity and academic achievement for subjects beyond the cutoff point in IQ, despite the fact that it is introduced in terms of this question.

More direct evidence was sought in another study by Yamamoto (1962), in which, hypothetically, a group of high creative and one of low creative individuals were to be compared on achievement, with intelligence level held constant. In fact, however, there was a substantial difference in the IQ levels of the two groups, suggesting that creativity and intelligence scores were too highly correlated to make possible an actual equation of the groups on intelligence. Yamamoto reports that,

with IQ held constant statistically, creativity scores were positively correlated with achievement indexes. Such statistical control is an unacceptable procedure in this case, since it assumes a linear relation between intelligence and achievement, and hence "begs the question." It is particularly inappropriate in that the mean IQ for the low creative group was below 120, while that for the high creative group was above 120; the notion being tested explicitly assumes that the relationship between IQ and achievement is not linear in such a case.

It appears, then, that Torrance's creativity measures are not independent of IQ in groups of average and slightly superior intelligence, and that they do not correlate strongly with one another. There is some evidence that, in intellectually superior populations, IQ does not differentiate on academic achievement. No demonstration has been provided that creativity is positively correlated with achievement for average or for superior subjects. Finally, the postulated correlation between achievement and "creativity," if it were to appear, would not conclude the case for creativity determination, since academic achievement may be regarded as an inappropriate criterion against which to judge indexes of creativity.

Mednick. The research described above gives no strong indication that a dimension separable from intelligence and appropriately labeled "creativity" can be isolated. The work to be discussed in this and the following section, on the other hand, does suggest that it is possible to find such a dimension. Further, this work offers a number of leads as to the correlates

of this dimension.

Mednick defines the creative thinking process as "the forming of associative elements into new combinations which either meet specified requirements or are in some way useful." (1962, p. 221) He notes further that "the creativeness of a product is some function of the number of requirements that the product meets." (p. 226)

His Remote Associates Test (RAT) provides the subject with a set of three distinct associative elements (words), and requires that he produce a word which will provide a link between the three--that is, a word which has an associative link to each of the three. A sample item would be the presentation of "rat," "blue," and "cottage;" the word "cheese" is the desired product, since it is associatively related to each of the three words given. In constructing the test, care was taken to ensure that the mediating link be "strictly associative rather than being of a sort that follows elaborate rules of logic, concept formation, or problem solving." (p. 227)

The RAT is given as a test. The subject is allowed forty minutes in which to complete the thirty-item instrument; hence, time pressure should not be great. Wallach and Kogan offer the criticism that, as with tests of convergent thinking, there is for each item a single (or occasionally several) right answers; thus, so far as the experimenter is concerned, the subject cannot produce any answer that is both correct and surprising. Thus creativity is measured by remoteness rather than novelty of product. This point may not be of great importance in itself;

however, it does imply that an upper bound on the possible demonstrated creativity is imposed by the test. This bound is that resulting from limits in the experimenter's creativity in formulating test items. Subjects capable of producing connections that the experimenter was not aware of have no opportunity to demonstrate this capability. This criticism, however, loses its force if the RAT does provide valid differentiation of individuals.

Mednick (1962) reports some research attempting to establish the validity of RAT scores and their independence of convergent intelligence indicators in selected (college) populations. In several college populations, faculty members who had good knowledge of the capabilities of students scoring at the extremes on the RAT rated their students in terms of creativity. Ratings differentiated high from low RAT scorers in two cases. In one of these, Miller Analogies Test scores were available on the students; these did not differentiate highs from lows on the RAT. In several undergraduate populations, a negative correlation between RAT scores and grade-point average, on the order of -.3, was obtained. There are differences between high and low RAT scorers on the Strong Vocational Interest Blank, with highs preferring vocations like artist, psychologist, author; and lows preferring vocations like farmer, teacher, pharmacist.

RAT scores of 38 college students correlated positively with the number of words produced in a one-minute period of writing as many associates as possible to each of 20 words (Mednick, 1962). Both quantity and uniqueness of anagrams

produced in a five minute period from the word "generation" were related to RAT scores.

These studies provide some limited support for the RAT as a means of separating individuals on a basis which may be independent of standard intelligence and achievement indexes (though why correlations with achievement should be negative is not clear) and whose correlates have some face validity as indexes of creativeness.

Mednick's theory invokes the notion of the associative hierarchy. Each associative element (i.e., word) in the individual's repertoire has linked to it a number of other associative elements. The "strengths" of the associative links (e.g., as measured by probability of evocation in a free association task) are varied. The elements connected with a particular element, in order of the strength of their link, form the associative hierarchy for that element.

Of particular interest for the study of creativity are differences in the characteristics of these hierarchies. Some individuals may be characterized as having steep hierarchies, in which the strongest associates have very strong links with the master word, while remaining associates have weak links. Others have flatter hierarchies; the strengths of the dominant associates (which may be the same associates as in the first case) are less than for the steep hierarchy, but there are more associates with intermediate-strength links to the element in question. More creative individuals are those with flatter associative hierarchies, since they have greater possibilities of coming to distant

not readily available connections among elements.

Several interesting results follow from the hierarchy notions. Individuals who are able to produce a large number of associates should also produce more that are uncommon. Their first associates should be the common, high-strength associates which are strong in all individuals; but after these, they have remaining some that are less common and which, in individuals with steep associative hierarchies, have low or zero strength. Within the individual, then, average uncommonness of response should increase as successive responses are given. Several studies already mentioned support the latter prediction.

Since latency is also a measure of response strength, some additional predictions are possible. The most common responses should be given with shortest latency, both within the individual and for the group. Thus, for the individual, rate of production should decrease with time in the task, even when the number of possible responses is very large. Also, in Mednick's formulation, latency for dominant responses should be longer in individuals who produce many associates than in those who produce few, since response strength of the strongest associates is hypothesized to be weaker in individuals with relatively flat gradients than in those with steep gradients.

Creative solutions are seen to occur when the requisite elements are evoked contiguously. This may occur by accident, by "similarity" (undefined) of associative elements; or through the mediation of common elements. At this point the formulation appears to be deficient in control mechanisms. Two problems in

particular arise. We know that search through the associative hierarchies is not random, or guided solely by the set of strengths of associations that occur (Ach, 1905). Some scanning mechanism which imposes constraints on the train of associations is needed. At the other end of the process, an evaluational procedure is necessary so that search is terminated when the creative solution has indeed been achieved. Thus, the associative hierarchy provides only the substrate, rather than an adequate mechanism, for the creative process.

Wallach and Kogan. The fourth program of research to be reviewed is that of Wallach and Kogan (1965). These investigators, noting the diversity of tests previously employed in studying creativity, devised a set of procedures in which emphasis was upon fairly simple associative productions. They measured both fluency of response--total number of responses appropriate to the task; and uniqueness--number of appropriate responses given by only one subject in the sample. Each of the choices appears reasonable in the light of the numerous introspective reports which emphasize associative freedom (lack of critical pressure; openness to all forms of ideation; a playful attitude toward possible ideational combinations) as basic to the creative moments of outstanding individuals. Fluency and uniqueness should be highly correlated, if creativity does indeed form a unitary dimension of individual difference; this result has already been predicted in Mednick's formulation.

Five procedures were used. In the first of these, the subject was required "to generate possible instances of a verbally

specified class concept, such as 'round things,' or 'things that move on wheels.'" In the second, a common object such as a newspaper, shoe, or button was given, and the subject was to suggest possible uses for the object. The third verbal procedure required proposing bases of similarity between two objects, such as potato and carrot, or cat and mouse. For the remaining two procedures, visual patterns rather than words formed the stimulus material. The subject was to think of possible interpretations or meanings for various abstract visual patterns, in one case, and for abstract line forms in the other.

These procedures are similar to some of Guilford's, particularly to those of Associational and Ideational Fluency. They involve both figural and verbal stimulus material, which Guilford has proposed will lead to separate factors,but have in common verbal modes of response. The distinctiveness from Guilford's work arises not so much from the materials used as from the context in which they were administered. Here, an attempt was made to avoid any suggestion that the creativity assessment devices were "tests" or that evaluation of any sort was being made. Ss were given a carefully maintained game-playing atmosphere in which to work. They were tested individually, and without time limits. None of the standard intelligence tests were administered until creativity assessment had been completed. Thus, every effort was made to enhance subjects' feeling of freedom from assessment pressures, and to provide opportunity for playfulness and spontaneity in dealing with the creativity procedures.

The subjects employed were 151 children, the entire fifth

grade population of a suburban public school system in a middle class region, and somewhat above the general population norms in intelligence. Ten measures of convergent intelligence were used. Three were individually administered: the WISC Vocabulary, Picture Arrangement, and Block Design subtests (Wechsler, 1949). The remaining group administered procedures involved appropriate-level standardized tests of verbal and quantitative aptitude (The School and College Ability Tests, Cooperative Test Division, 1957a, b); and of academic achievement (Sequential Tests of Educational Progress, Cooperative Test Division, 1957c, d, 1959). The latter were included on the assumption that conventional tests of academic aptitude and of academic achievement have the major part of their variance in common.

By the criteria discussed above, Wallach and Kogan succeeded in determining a dimension distinguishable from IQ. For all 151 subjects, the ten creativity indexes showed substantial positive intercorrelations, ranging from .07 to .74, with a mean of .41.[1] Forty-one of forty-five correlations were significant at the .01 level. The instances naming test had slightly lower correlations with other tests than the latter had with each other--a range of .07 to .45, as compared with one of .20 to .74--but all tests had sufficient intercorrelation to justify the claim that they define a single dimension. Total and uniqueness scores gave no

---

[1] The correlations reported employed a total fluency score from which the total of unique responses had been subtracted to avoid statistical confounding of the two measures derived from a given test. Correlations involving uncorrected totals will generally be higher than those presented.

indication of systematic differences in the pattern of correlations; nor was there a separation between verbal and figural procedures.

One criterion, that of substantial correlation among tests of creativity, was thus met. The other criteria proposed were also met. Aptitude and achievement tests showed positive correlations on the order of .5, with no separation of aptitude tests from those of achievement, or of individually-administered tests from those given in groups. Finally, correlations between creativity and intelligence tests were low, on the order of .1.

There were no sex differences in the pattern of results. For the 81 girls in the sample, correlations within creativity averaged .50; those within intelligence averaged .55; and the mean correlations between domains was .13. For the 70 boys, the mean correlations were, respectively, .34, .50, and .07.

The statistical separation of creativity from intelligence is thus quite clear in this investigation. It is particularly striking that the two programs which have succeeded in such a separation (this work, and Mednick's) have used creativity measures which require only relatively simple associative productions, and no particular skills in expression of ideas. They have also involved little time pressure--no time limits have been imposed by Wallach and Kogan; while those of Mednick's RAT are rather liberal.

An additional feature of their procedure which Wallach and Kogan believe to have been important was the absence of emphasis on evaluation of the subject, with an effort to present the creativity procedures as games. Support for this belief is given

by the failure to find a clear discrimination from intelligence of Guilford's Ideational Fluency and Associational Fluency factors, whose tests bear strong resemblance to Wallach's. On the other hand, Mednick's results indicate that the testing context does not make assessment of creativity impossible. The boundary conditions here are uncertain--note that Mednick's results were obtained with adults rather than children, and with a "test" bearing little similarity to intelligence tests.

Beyond their initial demonstration of the isolation of a dimension, Wallach and Kogan proceeded to investigate the correlates of status on the creativity and intelligence variables, considered jointly. For these analyses, individuals were ranked on the basis of summed standard scores for creativity measures and for those of intelligence. Independent median splits on each of the distributions were used to constitute four groups of subjects--those high on both creativity and intelligence; those low on both variables; and two groups of subjects high on one variable but not on the other. Grouping was performed separately for each sex.

The two by two classification on creativity and intelligence was used in analyses of variance of scores on the additional domains investigated. These involved the following: ratings of overt classroom behavior; self-rated anxiety and defensiveness; categorizing and conceptualizing; and expressive sensitivity. Additionally, clinical profiles of children in the various groups were drawn.

The various sources of information combined to present

a coherent picture of the ability and motivational syndromes of the four groups of subjects. The portrait which emerges is similar for girls and boys, although the sexes differ as to which procedures provide the relevant group differences. It will be useful to review these portraits, as they provide the best leads available for the study of creativity in children.

In the various analyses performed, it is clear that children high on both creativity and intelligence, and those low on both variables, are not at opposite extremes, with other subgroups falling midway between. Rather, while those high on both variables are in general superior to all other groups, each of the subgroups which are high on one variable and low on the other appear in some cases to show the least desirable attributes of all the groups. The kind of behavior on which members of the group show less desirable performance is, however, different for the two groups.

Those who are high on conventional intelligence but low on creativity give the appearance of over-adaptation to the norms of the academic setting. They show a lack of spontaneity, overcontrol, and high conformity to teacher expectations. The clinical descriptions for both sexes show an apparent self confidence and high capability for classroom performance; but suggest that an underlying fear of failure may lead to compulsive investment in academic excellence and in the meeting of expectations. Overt classroom behavior shows girls high in intelligence but low in creativity to avoid social contacts; to have undue hesitancy in self expression; and to avoid any kind of

disruption of classroom activities. In contrast to children high on both creativity and intelligence, who also show high involvement in intellectual activities, these children appear to receive little joy from their intellectual achievements; rather, these achievements may serve as a defense against doubts as to self-worth, so that good performance functions to prevent anxiety rather than to provide pleasure or self-fulfillment.

The inhibitions of this group appear strongly in some of the procedures employed. An object-sorting task shows these boys to be unwilling to venture beyond groupings of objects which meet the conventional requirements of "mature," abstract categorizing. While other procedures show that they have the capability to integrate elements thematically when so directed, in a free situation they use only conventional groupings. The groups high on creativity, by contrast, use both thematic groupings and conventional, abstract ones. This group seems to be motivationally blocked against venturing beyond the kind of response typically called for in testing situations, giving only responses which are sure to be "right" rather than ones which may be "good" though unconventional.

For girls high on intelligence and low on creativity, a similar lack of adventuresomeness appears in an expressive-sensitivity procedure, where the child is required to accept or reject an emotional state as possibly appropriate to the pose taken by a stick figure. Ss of this group accept only those emotional states which are highly appropriate, rejecting all those which are not clearly appropriate. The high creativity

groups, on the other hand, show the same acceptance of appropriate emotional states but are also willing to accept some not-so-clearly correct attributions as potentially appropriate.

Various kinds of indicators, then, show this group to avoid the unconventional and the possibly "wrong," and to have chosen a "safe" way of meeting the world. Conformity to classroom demands assures acceptance and recognition, and perhaps reassurance to the child of his own self worth. Control is too tight, however, for in situations where external evaluational pressures are relaxed, performance is unchanged. The child does not adopt the playful, spontaneous approach that the experimental procedures here attempted to make possible; nor does he indulge in divergent thinking even when he can do so in addition to, rather than as a substitute for, more conventional forms of behavior.

This group stands in sharp contrast to the children high in both creativity and intelligence. These show the same ability level, by conventional criteria, but show in addition a lack of subjugation by externally imposed standards. When it is appropriate, as in the procedures of this study, to abandon the convergent thinking atmosphere of the classroom, the high-high group is able to do so. Whereas the other group showed over-control, the high-high group shows adequate but flexible control. These children sort objects in terms of abstract criteria, but also thematically; they can determine appropriate affective labels for stick figures, but will also accept some prospective labels which are highly unusual. They show interest and

involvement in academic activities, but can become disruptive in the classroom when activities there are not to their liking. Girls in the group high on both creativity and intelligence show high levels of self confidence and little hesitation in self expression. Sociometrically they are interested in others and have this interest reciprocated. The portrait of this group is one of autonomy and self-direction, with the ability to function in widely diverse spheres, and a lack of motivational inhibitions of such functioning.

Children low on both creativity and intelligence generally give low-quality performance in all spheres, as might be expected. Their reactions to inadequacy, however, are varied. Among the boys, some individuals manifest a hyperactive "acting out," while others show general passivity. The girls show tenseness and insecurity in overt behavior, while self-reported anxiety is high among the boys. No mature efforts to advance status through work activities are shown by either sex. However, there are some cases of good compensation; among the girls, some interest and involvement in social activities is evident.

The final group, that of children high on creativity and low on conventional intelligence indexes, is the most difficult to pigeonhole. These children possess greater resources than those low on both variables, but their apparently greater sensitivity to themselves and to their environment appears to be a mixed blessing. The clinical description indicates that they are "children in angry conflict with themselves and with their school environment, beset by feelings of unworthiness and

inadequacy." Girls in this group, for example, show low attention span and concentration in academic settings; lack of confidence in self-expression; an avoidance of social contacts; and a high incidence of disruptive attention seeking. In a permissive setting, with ample encouragement from an adult and a lack of evaluational pressure, these children show flexibility and spontaneity in dealing with unfamiliar material; but when these conditions are not present, they are unable to cope adequately. Feelings of inadequacy, which the high intelligence-low creativity group has avoided by an addictive concentration on academic conformity and superiority, are much closer to the surface. In the clinical descriptions of the girls, there is an indication of overt passivity; a rich fantasy life may substitute for mature dealings with the world. In the boys, on the contrary, a note of impulsivity appears. Violent physical outbursts, disruption of class activities, and a short attention span suggest that impulse control is inadequate.

Some Efforts at Interpretation

Further Discussion of Wallach and Kogan's Results. Wallach and Kogan provide the most differentiated analysis available of creativity studied at the trait or dispositional level. Further discussion will center around their work, with Mednick's formulation considered as it becomes relevant. This section is devoted to a brief discussion of Wallach and Kogan's results. In

the one following, an attempt will be made to map their interpretations into another language system.

Wallach and Kogan make a first distinction between individuals with ability and those without. Those low on both creativity and intelligence are presumed to suffer a general ability lack. Support for this interpretation is provided by the relatively poor performance shown by these individuals on virtually all kinds of measures.

For the other groups, ability lack is not a reasonable interpretation; rather, each of these groups shows good performance in at least one of the two assessment contexts employed. Deficiency appears to be situation-specific. Wallach and Kogan attempt to explain the failure of each high-low group to score high-high; they focus this explanation on a particular motivational syndrome for each. The nature of this syndrome is clear in the descriptions already given. Take the group of children who are categorized as high on intelligence and low on creativity. These children seem to show overinvestment in academic achievement, with a consequent rigidification of behavior along the lines demanded by the norms of the classroom. They are thus highly "conservative." They carry the standards of ability-testing situations with them, and do not relax these standards even when the situation should suggest such a relaxation. Their thinking seems to be limited to conventional lines which reduce the possibility that responses can be judged wrong or inferior. There is the suggestion that they are threatened, not by a situation in which adequacy is being tested--since their characteristic

mode of behavior assures success here--but by an unfamiliar situation, in which norms are not made clear. Finally, they may be described as showing ego involvement, rather than involvement in the task.

Children high on creativity and low on intelligence, on the other hand, show good performance under just the opposite conditions. When personal adequacy is not stressed, they show strong task involvement. Wallach and Kogan suggest, in fact, that high status on creativity may entail a kind of compulsive involvement in the task, with consideration of many alternatives and an unwillingness to relinquish a problem until all sides of it have been examined. These children, however, cannot tolerate the stress of a competence-assessment situation; as soon as attention is focused on adequacy of performance, rather than on the task itself, they are unable to cope with the situation. Inadequate ego strength and a fear of failure, must then enter into an interpretation of their performance.

The final group, that of children high on both creativity and intelligence, is characterized as having the best points of each of these two. These children have sufficient resources not to be overcome by the pressure of ability evaluation; nor by the ambiguity and uncertainty of the creativity-assessment situation. They show maximum flexibility, following the lead of the situation rather than of any self-imposed limitations brought from outside.

An entirely different sort of analysis is provided by Mednick's discussion of associative hierarchies. Mednick

proposes bases for difference between creative and non-creative individuals, without considering differences in intelligence level; but his theory could easily be extended to embrace four, rather than two groups. Creative individuals are presumed to be those whose associative hierarchies include many associates of moderate strength beyond the dominant associates which are available to everyone. To extend this analysis, several classes of associates may be considered. One class would include conventional, logical associates, of the sort that should be appropriate in problem solving situations and for academic achievement in general. Individuals with many such associates represented at high strength in their hierarchies would perform well on intelligence and achievement indexes. Another class would be that of associates following less conventional lines, whose relationships might not be logical, but which nonetheless could be meaningfully related to their stimulus words. These would be useful in situations like that of creativity assessment. Finally, there would be a wastebasket category, comprised of associates fitting neither of these categories. These would include the various kinds of associates generally considered immature--e.g., clang responses, affective ones, and most syntagmatic responses (Flavell, Draguns, Feinberg, and Budin, 1958). Individuals with many such associates would perform poorly on both creativity and intelligence procedures.

While an extension like this could be made easily, there is a basic implausibility in Mednick's formulation. Mednick proposes that creative subjects "have" the requisite associative

connections, while others do not "have" them, at any appreciable strength. If strength is defined in terms of probability of evocation, then Mednick's proposal is truistic. Barring this, another criterion is just as reasonable.

Consider the example given from the Remote Associates Test, in which the subject must produce _cheese_ as an associate to the triad _blue_, _cottage_, and _rat_. No English-speaking college student should be completely lacking in knowledge of the existence of blue cheese, of cottage cheese, or of the proverbial love of rats for cheese. If the criterion were one of recognition or of acceptance of the associative link, rather than of its production, the creative response should be within the repertoire of virtually all subjects. It follows, then, that the question is one of why individuals differ in the associates they _produce_, not of differences in the connections they _have_.

Suggestions for a Model. It may be useful to consider these differences in the terms of another language system, that of "cognitive controls." This is model building at the level suggested by the plans of Miller, Galanter, and Pribram (1960); or by the information-processing routines of Newell, Shaw, and Simon (1958, 1962). Two reservations are necessary at the outset. First, I am presenting not a model, but only a mechanism which would be an important part of such a model if it were to be constructed. The effort is to account for the differences among groups, rather than to provide a complete specification of structure. Second, these suggestions are offered, not as an alternative to the interpretations made by Wallach and Kogan,

but as a reformulation of these interpretations.

As a basis for the differences obtained, consider a "cognitive control" which serves as a monitor on the processing and output of associative products. In a computer program, such a monitor would be a routine which would come into operation after the imposition of a problem and the setting up of its conditions. This routine would have two tasks: (1) Specify the thoroughness of search for solution; or, stated otherwise, specify the minimum degree of match between product and requirements which will be acceptable; (2) Evaluate potential products for acceptability. A product may be accepted, and output; or it may be returned for further processing; or, finally, it may be rejected altogether.

A more differentiated model might eventually require two such controls: one to govern generation and evaluation of steps toward problem solution; and one to examine potential products, to determine whether or not they should be output. Until such differentiation is required by data, however, and until independent measures can be devised for two such monitors, it is appropriate to consider only a single mechanism.

Several parameters must be specified for the operation of the mechanism. The first parameter is one of strictness of control. This refers to the thoroughness of processing, and the degree of match between product and problem requirements, required before a product will be accepted and output. As an example of the distinction intended, consider the case in which an individual is given as a problem, "Name objects which are small, round, and white." Suppose the potential product "beachball" occurs

to the subject. On the one hand, he may produce "beachball" as a response, with little consideration of the degree to which it meets task requirements. On the other, he may consider more carefully and/or demand a closer match between product and requirement, so that "beachball" is not given as a response; or perhaps the compromise, "small white beachball," will be given. In a problem like this one, different degrees of strictness can be seen in requiring, on the one hand, that the object named always or typically have the values demanded by the task; or on the other, that it may have these values among others--thus, for example, "leaf" is an appropriate item when naming red things, in the second case but not in the first.

The second parameter to be considered is one of flexibility of level of strictness. On the one hand, an individual may maintain rather inflexibly his characteristic strictness of control, without regard to the conditions of the problem situation. On the other, strictness may be varied as a function of the task, so that processing is thorough and cautious at some times, less thorough and more risky at others.

Finally, a parameter is needed for adequacy of control; here are implied differences in effectiveness under stress. When solution of a problem is highly important, for example, or when the recent history is one of repeated failure, control may break down, leading to failure or continued low-quality performance; or on the other hand, it may be resilient enough to withstand such stress.

These three parameters are sufficient to characterize

the differences in control among the four groups isolated by Wallach and Kogan. First, take the group high on intelligence and low on creativity. Good performance in an academic setting, with strong involvement in intellectual achievement, suggests that control is good in terms of strictness and adequacy. Standards are high, and they are not disrupted by the stress of academic assessment. However, these children show little flexibility: When evaluational pressures relax, they do not; they are unable to adopt a spontaneous approach or to relax concern with correctness of response.

Children high on both creativity and intelligence show the same high level of strictness and adequacy in their academic performance. They also show greater flexibility of standards. In a permissive setting, they can and do adapt to the change in what behavior is appropriate.

Children low on both creativity and intelligence are harder to characterize. Wallach and Kogan set them off from the other three groups as showing an ability lack. Such a statement, however, is imprecise. It might refer to a deficiency in repertoire: for example, to a lack of appropriate subroutines for information processing. On the other hand, it might refer to a failure of control: inability to set appropriate criteria for performance; failure to shift criteria in accord with task demands; or perhaps breakdown of control in the face of stress, where stress may arise from the pressure of academic evaluation, or from the unfamiliarity of the creativity assessment situation. No definite placement is possible for this group, since any or

all of four different deficiencies may be present.

Finally, there is the group of children high on creativity and low on intelligence. In a permissive setting, these individuals manifest strong task involvement and spontaneity in dealing with unfamiliar material. Here, then, sufficiently strict control appears. However, control is inflexible and inadequate: under the pressure of academic assessment, in which monitoring must be highly strict, control breaks down. These individuals are unable to screen out inappropriate responses and to output just those that meet task requirements.

This analysis does not contradict that made by Wallach and Kogan; its usefulness must be defended on grounds other than that of providing predictions contrary to those they might make. A minor value is that it points to an imprecision in their discussion: "Ability lack" is not an adequate description for children low on both creativity and intelligence, since it may denote any combination of a number of more specific failings. The language used here is suitable for a more differentiated treatment of such a lack, when data become available.

More important, this analysis provides a start for the study of process at a level closer to that of mechanism than that adopted by Wallach and Kogan. It is too incomplete to be called a process model, but is a step in the direction of such a model.

Finally, the analysis suggests some relationships between the analysis of creativity and intelligence, on the one hand; and Kagan's reflection--impulsivity dimension, on the other. As will be seen, substantially the same predictions can be made in

the terms in which Wallach and Kogan discussed their results; but the predictions seem somewhat more straightforward when derived from the present formulation, and were in part suggested by it.

In the section following, Kagan's research will be described in some detail. Subsequent discussion will consider relationships between this research and that by Wallach and Kogan.

Reflection--Impulsivity

Kagan's work began with an interest in concepts used in grouping objects. The Conceptual Style Test (CST) was designed to elicit two kinds of concepts: "Analytic" or physicalistic-descriptive concepts, which are formed on the basis of overt objective similarities of objects or of their parts; and "relational" concepts, in which the various objects in a group are related thematically, and each item alone is not an independent instance of the concept. An example of the first kind would be a group embracing a zebra and a striped shirt because they both have stripes; while in the second type of concept a pipe and matches might be linked, because the matches are used to light the pipe. A third kind of concept is the categorical--inferential, in which an abstraction superordinate to the members of the group forms the basis for classification. The CST was designed to suppress the frequency of grouping on this basis.

Versions of the test were used with adults, but most of the experimental work has been conducted with children in the

elementary grades. While a number of interesting correlates of relative preference for analytic concepts have been found, the analytic style is now seen as a composite variable chiefly attributable to scores on two more basic dimensions. One of these, visual analysis, concerns the ability or tendency to analyze into elements a relatively complex visual input, and is measured by transfer of learned names from a visual stimulus pattern to the elements of which this pattern is composed.

The other dimension, reflection--impulsivity, is of chief interest here. It is defined in terms of latency of response when there is some degree of response uncertainty. Thus, it is a decision-time variable, reflecting differences in the degree to which an individual "reflects" (inhibits impulsive responding) when there is uncertainty as to which of several simultaneously available response alternatives should be chosen. The alternatives may be either externally presented ones, or internally generated possibilities.

In conceptualizing this variable, Kagan (Kagan, Rosman, Day, and Phillips, 1964) points out that age differences in cognitive performance, as well as individual differences within an age range, are most often conceptualized in terms of differences in the availability and structure of transformation systems--that is, of relatively fixed "routines" or "schemata" for information processing. Such transformations, however, represent only one of several stages of information processing in problem solving or concept forming. Kagan points to three stages that must be considered--initial encoding of information, storage in

memory, and transformations. A fourth must be added, as he recognizes elsewhere but does not make explicit here: that of evaluation and output of cognitive products. Kagan suggests that stages other than that of transformations may show age and individual differences. He sees visual analysis as a dimension of particular importance at the "input" end, dealing with initial encoding of stimulus material. From his discussion, reflection--impulsivity should be seen as an "output" variable, dealing with the rigor of the individual's evaluation of products before committing himself to one of several alternatives. Kagan is willing, however, to consider extension of this concept, and to recognize that two aspects of reflection--impulsivity may be differentiable: the one concerned with rigor of evaluation of products for output; and the other having to do with analogous differences in rigor of evaluation of subproducts during internal processing (Kagan, personal discussion). In either form, the dimension seems quite germane to the earlier discussion of control principles in creativity and intelligence.

Kagan's tests for the measurement of reflection--impulsivity require difficult perceptual discriminations. Three tests have been used, with populations ranging from first to fourth graders. In the Design Recall Test (DRT) the subject is first shown a card on which a nonsense figure appears. The card is removed and, after fifteen seconds, presented again along with eleven variants, which differ from the original in size, orientation, curvature of lines, angles of lines, and other features. The subject must choose the figure that was originally presented.

Response time for first choice, and number of errors to correct choice, are measured for each of a number of figures. In the Matching Familiar Figures Test (MFF) meaningful figures are used, no memory element is involved, and there are six figures to be compared with each original; otherwise, the test format is identical to that of the DRT. Finally, in the Haptic--Visual Matching Test (HVM) the subject explores a cutout figure haptically until he is satisfied that he can identify it, and then must find the original from among a set of visually presented alternatives. Response time to first choice and number of errors are again the scores derived.

Sufficient evidence is available to support the claim that these tests measure a single dimension, and that this dimension has reasonable stability over time (Kagan et al.,1964). For example, in one study with third grade children, measures of response time on the three tests of reflection--impulsivity were all positively correlated; error scores were all positively correlated; and all latency scores correlated negatively with all error scores. Most coefficients fell within the range from .3 to .6. In another experiment, with second and third graders, latencies for DRT showed an average correlation of .56 across a nine week period. Including data from a single administration of CST, mean correlation was .49. The time and error scores for the DRT in this study showed correlations in the .40's with scores on the MFF test administered one year later.

In summarizing the results of a number of investigations, Kagan (1964) suggests that latency shows greater stability and

consistency across tests than do error scores, but that both are relatively reliable. Time and error scores generally show negative correlations to one another, providing one indication that the latency measures are not simply determined by differences in ability.

Error scores show negative correlations with verbal ability, with visual analysis, and with preference for analytic concepts on the CST. Response time measures, however, are relatively independent of visual analysis and of verbal ability. Most correlations with verbal ability (measured by several WISC subscales or by the CTMM) fall within the range from -.2 to .2, with median correlations around .1 (Kagan, 1964; Kagan _et al_., 1964).

While most subjects who show long latencies on the reflection--impulsivity procedures have low error scores, and most with short latencies have high error scores, two smaller divergent subgroups appear in Kagan's results. The first is composed of bright individuals who, on easier procedures such as the DRT, have short response times but make few errors; these, however, have longer response times on harder tests of impulsivity like MFF. Choice of an optimal difficulty level for the group being studied, then, is necessary; tests that are too easy hold little response uncertainty for the bright subject, and hence do not show whether he tends to prolong processing of information when uncertainty is high. A second small group "contains children whose long response times result from extreme fear...These _Ss_ fail to respond quickly because they have no idea of what to say and are afraid of offering any answer." (Kagan _et al_., 1964,

pp. 17-18)

There is no indication that age or sex affects the pattern of correlational results. In one study, however, third grade males made fewer errors on DRT than did third grade females, while the girls had longer latency on DRT and on CST. In another study, second grade girls made more errors on DRT than did boys, for each of two instructional sets; while response times showed no sex differences. (Kagan et al., 1964) These findings suggest that the regression of time on error scores will be different for the two sexes, though the magnitude of the correlation coefficient may not be.

With increasing age, error scores decrease, number of analytic concepts on the CST increases, and response time measures increase. There is the possibility, then, that separate regression coefficients will be needed for different age groups. The increase in reflection with age indicates that the improvement vis-a-vis error scores is probably not simply an improvement in "transformational" ability, but rather is due to more adequate scanning and checking of alternatives before responding.

The age increase in response latency appears also in a procedure in which a picture is presented tachistoscopically for the subject to identify. Anomalous pictures (e.g., a living room with a tree in the center) were used in one study (Kagan et al., 1964). Latency to first response was greater for eighth grade children than for third graders. Again, the development of a "control" function, rather than an increase in problem-solving ability, is suggested by the results.

In two of the four groups of second and third grade children examined in the same study, a "substantial" correlation between tachistoscopic response time and verbal ability appeared (while correlations between verbal skills and reflection--impulsivity measures of response time were on the order of .2). Kagan suggests that these two kinds of tests differ in that, for tests such as DRT, all response alternatives are simultaneously present; while, for the tachistoscopic procedure, the subject must generate the alternatives as well as choose among them. Verbally advanced children might be expected to generage more alternative responses, so that two factors would lead to longer latencies: the additional time needed for the generation of many alternatives; and the time needed for choosing among a larger number of possible responses.

It is important to note that tests on which there are not a number of simultaneously available alternative solutions do not lead to latencies correlated with those on Kagan's tests. A similar result was obtained by Thurstone in his factorial studies of perception (1944), in which speed was criterial for many of the scores. Three separate speed factors appeared, representing presumably more or less "central" processing. The "most peripheral" factor, comprised of tests of simple reaction time, was uncorrelated with the "most central" factor, Speed of Judgment, which was loaded with such tests as judging which of two cards had the more circles on it.

With some tests, such as the WISC Vocabulary scale, the correlation of latency with error scores might even be positive, or

U-shaped, rather than negative. Kagan's results apparently depend on the use of tests in which additional information or processing of information can increase the probability of a good response.

In view of these considerations, two rather different emphases in interpretation may be applied to individual differences in reflection--impulsivity. One emphasis would be more "cognitive," suggesting that reflective individuals are those who show relatively more mature forms of search for information in their problem solving activities. The increase in reflectiveness with age is consistent with this interpretation, as is the finding that, in the test of Haptic-Visual Matching, palpation time is as good a predictor of error scores as is latency of first response (Kagan et al., 1964).

On the other hand, a "motivational" emphasis would put chief stress on the impulse control required for reflective responding. Kagan suggests three possible bases for such differences in impulse control. First, anxiety over ability to perform adequately might make the impulsive child unable to tolerate the uncertainty of the period of silence and non-responsiveness needed for reflecting over alternative possibilities. He may fear that his withholding of response will be taken as a sign of inadequacy, and so blurt out the first hypothesis to occur to him. In support of the argument can be given the findings from a study in which children were required to learn a list of words. Impulsives gave more extra-list intrusions in recall than did reflectives. When the children were informed during the task

that their performance was inadequate, intrusions increased further for those classed as impulsive, while not increasing for those in the reflective group (Kagan, personal communication).

A second possibility is that high task involvement might lead the child who has some confidence in his ability to want to assure that his first response be correct. This is an explanation, then, of why reflectives respond slowly, rather than of why impulsives respond quickly. Indirect evidence comes from a finding that girls who reject traditional feminine interests are more reflective than those whose standards are more traditional (Kagan et al., 1964). It was shown earlier that degree of rejection of sex-typed activities by females is positively correlated with involvement in intellectual mastery (Kagan and Moss, 1962). Hence, a correlation of reflectiveness with involvement in intellectual activities is suggested.

Each of these possibilities puts some emphasis on the child's reaction to tests of his ability. A third alternative proposed by Kagan would make reflection--impulsivity a much more general aspect of the child's functioning. He suggests that either initial biological or early acquisitional differences in reactivity may be the basis for the particular differences which are observed in his tests. Early and subtle CNS damage is mentioned as a possible source, though general activity level apart from any brain damage is equally likely. In support, Kagan notes that: "The extremely impulsive boys in our studies were also excessively restless and distractible..." (Kagan et al., 1964, p. 38). He cites other evidence as follows:

With infants seen at eight weeks of age and again at thirteen months, motor activity and duration of sustained attention to external stimuli were stable over the period of 48 weeks, and were inversely correlated with one another. Extremely active one-year old infants were minimally attentive to intellectual problems at five and six years of age (Schaefer and Bayley, 1963). Ratings of hyperkinesis during ages three to six were inversely correlated with ratings of involvement in intellectual activity during adolescence and adulthood (Kagan and Moss, 1962). Kagan has also found a positive correlation in grade-school boys between cognitive impulsivity and proneness to taking physical risks.

Finally, the correlates of preference for analytic concepts may be cited as suggestive evidence, because of the positive relationship between reflectiveness and production of analytic concepts. Seven to ten year old boys who employed many analytic concepts "were less distractible in the classroom, less likely to display task irrelevant gross motor behavior in the playground or in a restrictive laboratory setting...They were more likely to become involved in sedentary tasks requiring long periods of concentration, more likely to prefer intellectual vocations that required motoric passivity (e.g., scientist, writer) and typically produced more complete drawings of objects." (Kagan, 1964, p. 13)

Some evidence, then, may be offered for each of four possible bases for the reflection--impulsivity dimension. These interpretations are not mutually exclusive; the dimension may

well have multiple determinants. It remains an open question whether, despite differences in the basis for scores obtained, the dimension is a meaningful one to employ. A given degree of impulsivity may well have different implications, depending on how it is arrived at. This seems particularly true if comparisons across ages are to be made; if changes by way of maturity are superimposed onto differences within age-homogeneous groups, many different combinations of the hypothesized elements could lead to the same placement on the dimension. Within an age group, however, the dimension could be meaningful even if its antecedents are diverse. Differences in thoroughness and effectiveness of information processing might have relatively uniform implications for the individual in a given age group, regardless of the developmental history of such differences.

Creativity and Impulsivity

The reflection--impulsivity dimension is interpreted in terms of tendency to inhibit first responses, in situations where further processing of information can increase the probability of a good response. The assumption is implicit that--except for one small group of deviants--further processing of information does occur in the interval through which response is withheld. On this interpretation, impulsivity seems to involve exactly the monitor on "output" of products that was hypothesized in discussing the mechanisms necessary to account for Wallach

and Kogan's four groups. That monitor was presumed to evaluate potential products to determine whether they were acceptable for disclosure to the world, whether they should be withheld temporarily, or whether they should be rejected entirely. More precisely, reflection--impulsivity can be identified as the level-of-strictness parameter of the monitor. The other parameters can then be measured by changes in impulsivity across task contexts.

A further basis for expecting relationship between the two sets of work under consideration lies in the similarity of explanations offered by the investigators. In each case, some stress is placed on task involvement, desire for achievement, and fear of failure as important contributors to performance.

There are then two viewpoints from which the relationships of these dimensions can be considered. First, they may be considered as coequal dimensions of individual difference in cognitive processes. The explanations proposed for each can then serve as a basis for predicting and investigating their intercorrelations. Secondly, tests of reflection--impulsivity may be proposed as measures of the output monitoring which has been hypothesized to explain some of the differences among creativity--intelligence subgroups. Emphasis is then on testing an interpretation of the creativity results by use of measures related theoretically to this interpretation. These perspectives may not be as clearly distinguishable in practice as they are on paper; nonetheless, it will be interesting to consider the predictions that can be made from each.

To begin with, consider reflection--impulsivity as a dimension of individual difference in its own right. The discussions given by Kagan, and by Wallach and Kogan, then suggest some relationships between this variable and that of creativity--intelligence.

The clearest of these suggestions is that the most reflective individuals should be those in the group high on intelligence indexes but low on creativity. Wallach and Kogan describe this group as showing cognitive cautiousness, unwillingness to risk error by deviating from conventional modes of response, high personal investment in academic achievement, and personal coldness and aloofness. Kagan, in discussing the main group of reflective individuals, characterizes them as concerned with intellectual mastery, manifesting high task involvement and a strong desire to be right on first response. Further, they are described as more analytic, more prone to solitary task concentration, and less likely to take physical risks. The overlap of descriptions is thus quite striking, and provides a strong basis for this first hypothesis.

For other subgroups, predictions can be made, though with somewhat less confidence. A second hypothesis to be considered is that among the more creative individuals will be found those who are most impulsive. More precise versions of such an hypothesis can take several forms. The most straightforward prediction would be that creative individuals are more impulsive than noncreatives, regardless of standing on intelligence assessment. Creativity might entail minimal censoring or evaluating of

potential responses before they are made public; and such minimal censoring would lead to relatively quick responding.

However, other factors must be considered. One possible interpretation for reflection--impulsivity involves the effects on intellectual performance of anxiety and fear of failure. The syndrome proposed for children high on creativity but low on intelligence invokes such anxiety and fear; hence, the hypothesis may be made that these children will show impulsivity.

Children high on both creativity and intelligence, on the other hand, do not show this syndrome. They are seen as free from the limitations on performance which are self-imposed by those in the high-low groups. Performance should depend on the task and on their ability to handle it; in particular, feedback as to adequacy or inadequacy of performance should lead to a realistic altering of response rate.

For any given test, then, it is not possible to predict where children high on both creativity and intelligence should fall. The other groups should show less flexibility across tasks. Those high on intelligence and low on creativity should respond slowly and cautiously in general; while those high on creativity and low on intelligence should typically show quick, impulsive response.

These results depend on use of procedures in which the assessment of adequacy is made salient. Kagan's methods involve this emphasis: the subject is told his response is wrong whenever he makes an error; and he must continue search for the correct response until he has made a number of errors. Suppose,

instead, that measures of reflection--impulsivity were obtained under permissive conditions, in which any choice made by the child would be accepted as a good response. This change would not be expected to alter performance greatly for those high on intelligence, since these children are not threatened by evaluation. It might, however, affect performance considerably for children high on creativity but low on intelligence. These individuals appear to be capable of high task involvement and good performance, so long as personal assessment is not at issue. Hence, there is less reason to expect the group differences hypothesized to appear under conditions of minimal evaluation. If the atmosphere is completely benign, and responses are accepted as adequate whether or not they are correct, only the child's generalization from other situations in which evaluation has been salient will prompt the appearance of the differences suggested. Some such transfer may be likely in any procedure, since the child who has learned to fear evaluation by adults is unlikely to become completely at ease when alone with a relatively unfamiliar adult; but the differences will be much less extreme than in a context which emphasizes adequacy of performance.

For the final group, children low on both creativity and intelligence, some basis exists for proposing that they will show impulsivity on Kagan's procedures. The characteristics of this group include a lack of intellectual involvement and of mature efforts at achievement of any sort; and these accord well with parts of the description of impulsive individuals.

On the other hand, Wallach and Kogan suggest that some

individuals in this group show rather extreme passivity and withdrawal; it is uncertain for them whether low involvement in test procedures will lead to prolonged hesitation and withholding of response, or to impulsively quick response. Possibly, both will be found. Further, the high level of anxiety reported by these subjects might indicate that here will be found Kagan's divergent group of individuals whose long latencies are due to anxiety and lack of alternative responses, rather than to consideration of many possibilities.

Two different kinds of reaction then seem plausible for subjects low on both creativity and intelligence: extremely long and extremely short latencies. The group will then be characterized by great variability.

These predictions result from consideration of the proposed antecedents of creativity and impulsivity, taking each as a dimension in its own right. Reflection--impulsivity tests may also be viewed as measures of the monitoring proposed in interpretation of the results of the creativity study. The same predictions follow, quite straightforwardly, on this interpretation. Their derivation will serve in part to summarize and reinforce those made above.

First, individuals high on intelligence but low on creativity have been characterized as showing strict and inflexible control. They should then make up the most reflective group. Further, they should show reflection regardless of task context--in particular, both when the situation emphasizes evaluation of response adequacy, and when ability assessment is not

emphasized.

Second, those high on both creativity and intelligence are described as showing strict but flexible control. In Kagan's procedures, then, they should respond reflectively: here, response adequacy is emphasized, and feedback is given which should encourage adjustment of processing time to a level commensurate with accurate performance. When feedback is favorable, however, and all responses are accepted as good ones, these children might show less reflectiveness than those high on intelligence but low on creativity--their level of performance can be adapted to task requirements, and is not fixed by limitations from within.

Third, children low on both creativity and intelligence are marked with an ability lack, whose precise nature is ambiguous in Wallach and Kogan's results. This lack, however, may involve both deficiency in transformation systems (to use Kagan's terminology), and deficiency in control mechanisms. The former lack would suggest that prolonged withholding of response may occur--here, not because of careful consideration of alternatives, but because of ineffective search for them. Failure of impulse control, on the contrary, would suggest that "impulsively" quick responses will be given. Since both kinds of disability may be present, the prediction is for great variability in latencies. Unlike the variability expected for children high on both variables, this inconsistency is not expected to be correlated with task difficulty and context. It may be looked for both in the individual and in the group; and there is no basis

for expecting permissive vs. nonpermissive contexts to influence its appearance.

Fourth, children high on creativity but low on intelligence show sufficiently strict control, but this control is inflexible and inadequate. Their failure cannot be one of transformational systems, since in a benevolent atmosphere performance is good. It is only when involvement centers not on the task, but on the individual's ability to perform it, that performance breaks down. Thus, when no stress is placed on evaluation of performance, these individuals should be reflective. If adequacy is stressed, control will break down, and impulsive responses will be obtained from these children.

Both ways of looking at the impulsivity dimension, then, lead to a set of predicted relationships with the creativity and intelligence groupings. The measures devised to investigate these relationships must be discussed; but before this, it is appropriate to consider some possible age differences in these variables between younger and older children.

Age Differences

Some of the research discussed by Kagan (supra, pp. 61-62) suggests that individual differences in reflection--impulsivity will appear in children of preschool age. This expectation is strongest when cognitive impulsivity is seen as an aspect of a generalized disposition which also includes motoric restlessness, short attention span, proneness to take physical risks, and lack

of intellectual involvement. However, the tests have not been used with kindergarten children; it remains possible that they will not succeed in differentiating children of this age. This could arise if the "transformation systems" necessary for performance in complex perceptual discriminations are inadequately developed in many preschool children; so that individual differences in latency would be more indicative of ability to perform the task than of cognitive style.

Turning to the creativity--intelligence distinction, little direct evidence is available concerning age differences in the relationships of these variables. Wallach and Kogan's separation of creativity from intelligence in fifth-grade children, and Mednick's similar results with college level populations, suggest some stability over a wide age range. Torrance has worked with children in the first grade, and claims good results with some of his tests at this age; but the published results do not present evidence on test correlations (Torrance, 1963). Bowers, using the Minnesota procedures, reports a correlation of .26 between divergent thinking scores and Stanford-Binet IQ's for 150 elementary school pupils (Bowers, 1960). Finally, Iscoe and Pierce-Jones (1964) find correlations on the order of .20 between the WISC subscales and fluency scores on the Unusual Uses Test, for lower class white and Negro children aged five through nine.

Only the study by Iscoe and Pierce-Jones has examined children as young as those of preschool age; and this study employed only a single test of creativity. Hence, direct evidence

as to the possibility of finding the creativity dimension at this age is almost completely lacking. Various indirect sources of hypotheses, however, deserve some consideration.

To begin, note that the creativity procedures are heavily verbal in nature; even those procedures involving figural stimuli call for verbal responses. Children of preschool age differ from those older in various quantitative aspects of language; vocabulary size and relative facility with certain grammatical forms are among the obvious differences. The possibility must be considered that there are also qualitative, or structural, differences present. Data from word association tests suggest such a difference: young children typically give associates which are "syntagmatic"--words that can follow the stimulus in a sentence; while the normative responses of older subjects are "paradigmatic"--words that can replace the stimulus word in many contexts (e.g., Woodrow and Lowell, 1916). Further, Brown and Berko (1960) have shown that the tendency to give paradigmatic associates is correlated with learning of English parts of speech in age-homogeneous samples of children.

If these isolated differences in task performance are indicative of a structural difference in language between younger and older children, then it is possible that the creativity dimension, with its verbal measurement procedures, will not appear as a dimension distinct from conventional IQ. The same possibility can be suggested as a derivative of the hypothesis of increasing mental "differentiation" with age (Witkin, Dyk, Faterson, Goodenough, and Karp, 1962; Harvey, Hunt, and

Shroder, 1961).

Changes in language are one aspect of a broad-scale cognitive reorganization occurring within the period from age four to seven. The nature and extent of this reorganization is nowhere adequately examined; evidence for some change within this period, however, comes from such diverse fields of investigation as to demand consideration. Piaget, for example, has proposed a transition at about age seven from "preoperational" to "operational" intelligence, with a concurrent decline in "egocentrism" of language and thought (e.g., Piaget, 1926, 1928; see also Vygotsky, 1962; Flavell, 1963). Over 100 experiments performed in this country to test parts of Piaget's theory have provided confirmation, at least for the broad outlines of his argument (Wallach, 1962). Psychoanalytic theory has suggested a transition from "primary process" to "secondary process" thought at about this period (Freud, 1959b). Luria and other Russian writers who have followed the pioneering work of Vygotsky (1962) consider developmental changes in the role of language in the control of behavior, with an apparent fruition of this development during the same period (Luria, 1961; Luria and Yudovich, 1959). Studies of verbal mediation in problem solving and in concept formation point to a changing role of language in intellectual tasks of quite distinct phenotype (e.g., Kendler, Kendler, and Wells, 1960; Reese, 1962). Longitudinal studies of intelligence, reviewed by Braine (1959), suggest a change in relative standings of individuals from before to after this age. Additional sources of evidence are summarized by White (1965).

The age of transition suggested differs somewhat according to the kind of data adduced, and may also depend upon the general level of intelligence of the group studied (Braine, 1959). That some change does occur seems a reasonable conclusion, though its precise nature is quite unclear. It raises the possibility that the creativity dimension, reflecting some presumably general aspects of the individual's cognitive processes, may not appear in children below the age of transition, even if it can be determined reliably for children who have passed through this period: preschool age children may have available only a single mode of thinking, rather than the two distinct modes found in older children.

There is no basis for prediction of any specific differences between younger and older children. The foregoing possibilities merely emphasize that replication of the creativity--intelligence distinction in children of preschool age is not at all assured. Whether or not the dimension can be found, study of children of this age will serve as a first step toward determining the extent and importance of the transition in cognitive functioning which occurs at about the beginning of the primary school years.

If the dimension does appear at this age, Wallach and Kogan's interpretations may require some revision. These interpretations emphasized fear of academic failure, in one group; and need for academic achievement, in another. For children not yet introduced to the stresses of the academic setting, a more generalized set of dispositions is needed. While it would not be difficult to make this generalization, it is important to

know whether the motivational syndromes which appear in fifth grade children are the result of a particular aspect of the child's socialization, that concerned with adaptation to the conventional school environment; or whether, on the other hand, they result from more general aspects of the cultural conditioning of the child. The latter would suggest that academic fears should be seen as a special case of dispositions much broader in scope.

Aims and Hypotheses of the Present Investigation

The experiment reported here was designed primarily as a strong test of the generality of Wallach and Kogan's creativity--intelligence distinction, and of Kagan's reflection--impulsivity dimension. Wallach and Kogan, working with fifth grade children, found a number of measures of creativity to be highly intercorrelated, and to have zero-order correlations with intelligence and achievement measures. Their interpretation of results stressed several variables, anxiety over achievement and involvement in intellectual activity, which may be particularly important in the context of the conventional classroom. To test the generality of these findings, several of their procedures were modified for use with kindergarten children--children not yet subjected to the socializing influences of the classroom. The children tested were chosen to represent a broader range of intelligence and socioeconomic class than those with whom Wallach and Kogan worked; the latter were of above average intelligence and predominately upper middle class in background, while in

the present study IQ ranged from low normal to superior, and many children were of lower middle class or at the upper end of the working class. It was hypothesized that the creativity dimension was of sufficient generality to appear despite the differences in age, intelligence, and background for this sample--that these children, like those studied by Wallach and Kogan, would show significant correlations among creativity measures and nonsignficant correlations between these measures and a measure of verbal intelligence.

An important aspect of the procedure was the effort made to control the salience of cues as to the evaluation of responses for the various measurement devices. Wallach and Kogan saw the introduction of a playful, permissive testing context as necessary for the separation of creativity from intelligence. An attempt was made to duplicate this atmosphere for the creativity testing. The appropriate control condition, that of measuring creativity in the atmosphere standardly employed for ability testing, was not used; but, in the light of the failure of creativity--intelligence separation frequently found in the literature, positive results would provide indirect support for the importance of the permissive atmosphere.

The study of reflection--impulsivity has also been conducted with children of school age, again with some emphasis on motivational variables related to the traditional values of the classroom. The measures employed have all been perceptual matching tasks. The test for generality here involved not only the choice of a previously-untapped subject population, but the employment

of three new measures which are phenotypically different from those employed by Kagan, while conforming to his definitional requirements for measures of reflection--impulsivity. It was hypothesized that these new measures would show significant correlations with two measures similar to those used by Kagan.

Reflection--impulsivity was measured under two distinct conditions of salience of evaluational cues. For three procedures, an attempt was made to continue the permissive, game-like atmosphere used in creativity measurement; for the remaining two, the adequacy or inadequacy of his performance was made as salient as possible for the child. This contrast is of some interest for the question of the generality of reflection--impulsivity, but was introduced primarily for its relevance to the second major aim of the study. Here, a number of predictions were made for relationships between the child's status on the creativity and intelligence variables, considered jointly, and his standing on reflection--impulsivity. Chief among these are the prediction that children high on creativity and low on intelligence will be highly impulsive, but only when evaluational cues are highly salient; and that children high on intelligence and low on creativity will be highly reflective, regardless of the context in which reflection is measured.

These predictions were derived in two ways. First, there is some overlap in explanatory variables, largely motivational ones, used by Wallach and Kogan on the one hand, and by Kagan on the other. Relationships were hypothesized where common antecedents were proposed for the two domains. Second, a restatement of Wallach and Kogan's results was made in the language of

"cognitive controls," and reflection--impulsivity was conceptualized as an operationalization of one parameter needed to specify the working of the control process. The remaining parameters were seen as measurable through differences in reflection--impulsivity across administration contexts. The various relationships predicted could then be derived from the model of control processes.

This part of the study is quite exploratory in nature, since it is not yet known that either dimension will appear under the present conditions. Further, the design of the experiment emphasized obtaining many measures on a relatively small number of subjects, in accord with the purpose of assessing breadth and reliability of several dimensions. A result is that, when breakdowns are made for investigating interactions of variables, sample sizes are too small to give more than suggestive results.

CHAPTER II

PROCEDURE

Subjects

Subjects were drawn from four kindergartens in Durham, North Carolina. All children in kindergarten classes in these schools participated in the experiment. However, data for the several non-white children in School A were not included in the analyses; and there was insufficient time before summer recess to complete testing of all children in School D. In the latter case, children whose testing could not be completed were allowed to "play games" from time to time. One potential subject was excluded for refusal to participate; otherwise, no child showed unwillingness.

Characteristics of the schools are varied. School A is a private school whose pupils are drawn mostly from professional families, many of them university connected. It includes the youngest children tested, with three boys and six girls under 60 months of age. Children attend for a full morning five days a week. The program is somewhat unusual, in that each child spends approximately one and one-quarter hours of each morning in an Americanized Montessori classroom, and a similar period of time, with a different teacher, in a more conventional

kindergarten program.

Schools B, C, and D have pupils of more diverse background. In School C, families cover the range from upper middle class to working class (e.g., truck driver); while in Schools B and D most families are of lower middle class. B and C are private, while D is church-related. Children remain for a full day in each of them, with most organized activity occurring in the morning, followed by lunch, an hour or more of napping, and free play in the afternoon. No children younger than 60 months are enrolled in the kindergarten classes of these schools.

Table III presents information on subject age, intelligence, and social class. Information is given separately by sex, overall and with a school by school breakdown. Age was measured to the nearest month at the time of administration of the fifth of nine procedures. The IQ and mental age (MA) measures were obtained from Forms A and B of the Peabody Picture Vocabulary Test. The index of social class is described below.

Mean age was 68.0 months for girls and 70.0 months for boys. IQ averaged 100.6 for girls and 105.4 for boys. For neither of these variables was the sex difference significant. A sex difference did appear for MA, however, where females averaged 70.6 months; and males, 76.5 months ($F = 4.46$, $df = 1/85$, $p < .05$).

Several differences among the schools may be noted. For each sex, the analysis of variance shows a significant age difference across schools. By Duncan's New Multiple Range Test, girls at School A were significantly younger than those at each

Table III

Subject Characteristics

Group	Sex	N	Mean Age	Mean IQ	Mean MA	Social Class					
						I	II	III	IV	V	Un
All Males		41	70.0	105.4	76.5	8	4	13	4	1	11
All Females		46	68.0	100.6	70.6	16	2	6	3	1	18
School A	M	9	65.2	115.0	80.4	6	1	0	0	0	2
	F	16	61.2	109.8	73.1	14	0	0	0	0	2
School B	M	8	72.3	103.3	76.5	0	0	3	3	0	2
	F	13	70.7	92.0	65.2	0	2	4	1	0	6
School C	M	10	70.3	105.6	77.3	2	2	4	0	1	1
	F	8	70.5	98.3	70.4	1	0	1	1	1	4
School D	M	14	71.4	100.2	73.4	0	1	6	1	0	6
	F	9	73.9	98.9	74.2	1	0	1	1	0	6

other school ($p < .01$); while boys at School A were younger than those at School B ($p < .05$).

No significant differences across schools were found for the measure of mental age. For IQ, males showed no differences, but a significant F was found for females. By Duncan's procedure, girls at School A were significantly brighter than those at the other schools (compared with School B, $p < .01$; with C and D, $p < .05$).

Finally, Table III presents the available data on socio-economic status. These data were obtained from a questionnaire to which 67% of the parents responded. The classification is that described by Hollingshead (1965), and is based on a weighted average of occupation and education of the head of household. As a guide to interpretation, a professional with education beyond the undergraduate level would fall into Class I. Public school teachers and proprietors of medium sized businesses, holding a college degree, would fall into Class II. Class III includes high school graduates in occupations like insurance agent, private secretary, and radio announcer. Skilled manual employees with partial high school training are found in Class IV. Finally, a semi-skilled worker with junior high school training falls into Class V.

As has been indicated, children at School A come from upper middle class homes; twenty of twenty-one parents responding fall into Class I. For each of the other schools, the modal class is III. Collapsing across sex and analyzing for subjects above, in, and below Class III, there is no significant difference in

relative frequency of class assignments for Schools B, C, and D ($X^2 = 2.84$, df = 4); while assignments for School A differ significantly from those for the remaining three ($X^2 = 30.74$, df = 2, $p < .001$).

Testing Conditions

The physical conditions under which testing was conducted were adequate but less than ideal. It was necessary to use whatever space the kindergarten could make available. This space was in daily use for purposes other than testing; so that it was not possible to standardize features of the room from school to school. In each case, however, there was provided a relatively distraction-free area whose features remained constant from day to day. This area included a table, two chairs, and storage space, with furniture selected for the comfort of the child rather than that of the experimenter.

The experimenter faced the child across the table, making no effort to hide the clipboard and stopwatch used in recording and timing responses. Children not infrequently asked the purpose of these instruments. Their questions were answered and from time to time they were allowed to play with the watch. No child seemed upset or distracted by this recording.

All testing was done individually, with one procedure to a session and, with few exceptions, one session to a day. In School A, at the suggestion of the teacher, a fixed rotation was followed. Each child, when his turn was completed, tapped the

child who would follow him. In the other schools order of testing was haphazard, but so far as possible each procedure was completed with all children before the next was introduced. No child was taken at a time when he was unwilling to come; this entailed only two or three refusals throughout the entire experiment.

With the exception of the final two procedures, meant to be somewhat stressful, every effort was made to make the testing comfortable, pleasant, and fun, and to minimize any anxiety over adequacy of performance. Steps taken to realize this goal included the following. First, the experimenter (the author in all cases) was present at the school on two days before testing began. He interacted with the children during free play on these occasions, including an hour or more of active involvement in informal outside play. The latter was highly effective in breaking down restraint; generally, by the end of an hour, he was under massive physical attack by subjects of both sexes. During these periods, the children were told that in a few days they would have anumber of opportunities to play some unusual games, and that they would receive prizes on each occasion.

Second, so far as possible, the procedures that were used were all presented as games, with no indication that responses were to be evaluated on any criteria of adequacy. (Exceptions to this rule are indicated in the section following, in which procedures are described in detail). In each session, the experimenter assured himself that the child was interested and at ease before introducing the task. Throughout the task, liberal

praise and encouragement were given.

Finally, at the conclusion of each session, the child received a prize. Toys retailing for five or ten cents were employed; they included, for example, a large balloon; a bag of marbles; a birthstone ring; a pencil sharpener; a magnifying glass; and a very shrill whistle. Often the child was given a choice of prizes; and in each case he was allowed to select the one he wanted from a "treasure chest" or display.

Procedures

Procedures employed fall into four categories: warm-up; creativity assessment; intelligence assessment; and measurement of reflection--impulsivity.

Warm-up. The first session, devoted to free drawing, was used to familiarize the child with the testing situation. After an exchange of pleasantries and when child and experimenter were both seated at the table, the child was instructed as follows:

> "Now, (child's name), we're going to be coming in here a lot of different times. And each time, we'll have something different to do, or some different game to play. For today, do you know what it is?...Today I'm going to give you a piece of paper, and three Magic Markers, and get you to draw me a picture. Will you do that? Whatever you want to draw a picture of..."

The child received a sheet of paper and three felt-tipped markers (red, black, and purple). As he drew, E used his stopwatch and made notes on a record sheet. When the drawing was completed, the instructions continued: "Very good. Now, that's all we're going to do today. Another day we'll come back, and

we'll have something different to do. One more thing. Every time you come back here with me, you win a prize. For today, your prize is a whistle. Which one do you like? OK. Now let's go and see who's next."

Creativity. The second, third, and fourth procedures were employed in creativity assessment. Each measure was a version of one used by Wallach and Kogan (1965), substituting items of appropriate difficulty level for five and six year old children.

The Pictures test is most closely related to Wallach and Kogan's Pattern Meanings instrument. Nine abstract patterns were drawn with a red "Magic Marker" on 4 x 6 white index cards. The patterns are reproduced in Figure 1. Those numbered 7 and 8 were taken from Wallach and Kogan, while the others are new. Pattern 1 served as a sample item. All were sealed in clear plastic, so that the child could handle them freely.

As with items in the other two creativity procedures, patterns were selected on grounds of expected appropriateness and pleasingness for kindergarten children. Some pretesting of all procedures was conducted in a fifth school, to ensure that suitable items had been chosen. No revision of the creativity measures was seen to be necessary after this pretesting.

The administration of the Pictures test will be described in detail, as it exemplifies the format for all three creativity procedures. Instructions began:

> "Do you remember what we did the last time you were here?...That's right; you drew a picture for me. Well, today, I have some pictures to show you. And here's the first one. What do you think that's a picture of?.." (If no response, then: What does it look like? What do you think it could be?")...Good! And what else does it look like?...Very good!"

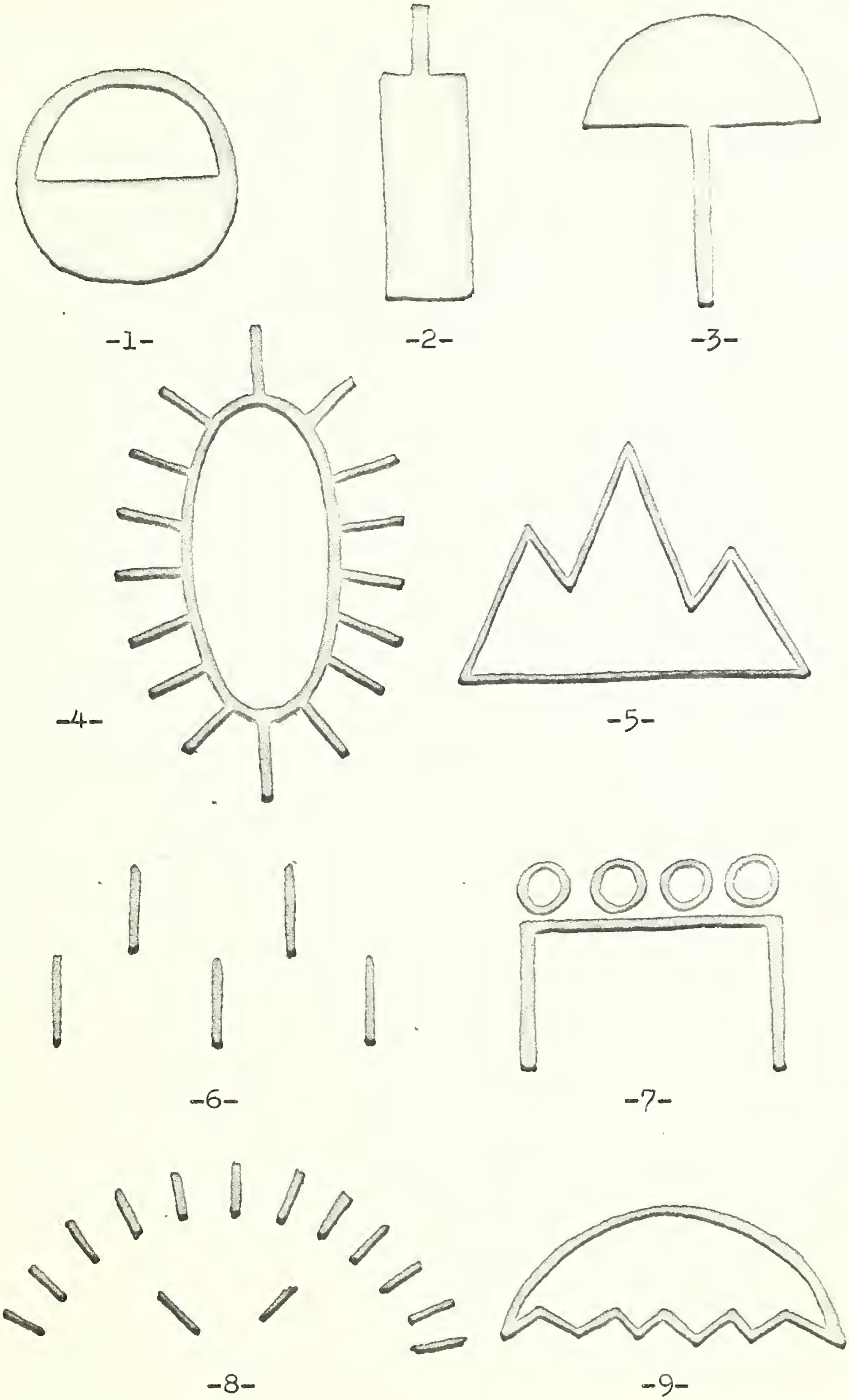

Fig. 1. Patterns for the Pictures Test

E continued prodding, and gave as examples of possible responses items from the set "basket," "pot," and "lady's pocketbook," until the child had given at least three plausible responses, or until he had otherwise shown--by accepting E's examples, by facial expression, by nodding his head, etc.--that he grasped the task as one of naming many possible interpretations, rather than of giving a single correct answer. A final example was then added, with: "Let me show you one more. Let's turn it around this way. Now, how about an army hat?" This last example was intended to indicate that the card could be looked at in any orientation. E added: "You see, there are lots of different things you can think of for these pictures. Now, let's try the next one. What can you think of for this one?"

E continued to use expressions like "good," "very good," "you had some good ideas for that one," throughout the procedure. When S seemed to have no more responses for a card, or to become distracted, E asked: "What else can you think of?" or a similar question. When no more responses could be elicited, the next pattern was introduced. Upon completion of the last card, additional praise was given.

The second creativity procedure was based on Wallach and Kogan's Instances test. The S was told: "Today we have a game with words. It's a game called 'Naming Things.' The first thing we'll do it with will be round things. How many things can you think of that are round?...What else is round?..."

The same general considerations apply to the Instances Test as to Pictures. When the child would name no more round things, he was asked in turn for "things that have wheels," and for

"things that are red."

Finally, a Uses test served as the third creativity measure. The items employed were, in order, a newspaper, a table knife, a cup, and a coat hanger. Instructions began:

> "Today we have a game called 'What can you use it for?' The first thing we're going to do it with will be a newspaper. (*E* hands a newspaper to *S*.) Now, I want you to tell me *all* the things you can *think* of that you can *do* with a newspaper, or *play* with it, or *make* with it. What can you use a newspaper for?..."

The Uses procedure was almost identical to that employed by Iscoe & Pierce-Jones (1964), whose *S*s included children of five and six years of age. Their fourth item was a clock, rather than a coathanger. Otherwise, except for the unknown factor of the testing atmosphere, this administration matched that described in their report.

In administering these procedures, *E* did not question the appropriateness of any response. If the child continued to give responses appropriate to a preceding item rather than to the item at hand, he was reminded what the present item was. If *E* could not understand the child's response, one request for a repetition was made. It was felt that asking for a second repetition would distract the child too greatly; therefore, responses still not understood were simply scored as "incomprehensible." Finally, on occasion a child was able to communicate the referent for a response without having available an appropriate word or phrase. In these cases *E* recorded the word intended.

Within these limitations, all responses were recorded verbatim. Also recorded were latency of first response, and the end of each successive ten-second period spent in responding to

the item.

Intelligence. The fifth and seventh sessions were devoted to Forms A and B of the Peabody Picture Vocabulary Test (PPVT; Dunn, 1959). The PPVT was chosen for several reasons. First, the test has a convenient format. The E presents a word, and asks the child to point to the one of four pictures which represents this word. It is ideal for use with young children, in that it provides interesting pictorial stimuli; can be given quickly; and requires no expressive skill on the part of the child, nor interpretation of response on the part of the examiner. All responses are accepted as good ones; the test is then an unstressful one, and can be maintained as a game without much difficulty.

Second, because of its format, the PPVT can be used to obtain a measure of reflection--impulsivity without added expenditure of research time.

Third, vocabulary has been held to be the best single indicator of intelligence (Wechsler, 1949; Terman and Merrill, 1937). If the distinction between verbal and performance ability is a valid one--and McNemar (1964) argues that in general it is not--verbal intelligence can appropriately be concentrated upon for comparison with the creativity procedures. These latter are heavily verbal in nature; hence, a demonstration of independence between a verbal intelligence measure and measures of creativity would be a particularly important finding. For a similar reason, it is appropriate to measure intelligence in as stress-free an atmosphere as possible: Possible attenuation of the relation-

ship between creativity and intelligence by differences in test context is thereby minimized.

Reflection--impulsivity. Three new measures of reflection--impulsivity were used. Two of these were obtained from the PPVT This instrument employs a series of items in order of increasing difficulty; the test continues until the child misses six of eight consecutive items. Latency of response on the last six errors provides an appropriate measure of impulsivity. Limiting measurement to these errors should equate difficulty level across children, yielding a measure which is uncontaminated by differences in ability to perform the task. Since Forms A and B of the PPVT are given to each child, there will be two impulsivity measures obtained on what are essentially alternative forms of the same test.

A Dots test was developed as another measure of reflection--impulsivity. This is similar to the Circles Test used by Thurstone (1944). The child is shown a 4 x 6 white card, divided by a vertical line at the center. Black ink splotches appear on each side of the card, and there are more of these splotches on one side than on the other. The task is explained as one of discovering and pointing to the side of the card that has more dots on it. After a sample, there are fifteen cards, each on a different page of a notebook. They are arranged, intuitively, in cycles of easy, intermediate, hard items; thus, the child is assured an easy discrimination at least once in every three trials. Latency of response on the five hard items forms the measure of reflection--impulsivity.

Finally, two Matching Figures tests are similar to Kagan's Matching Familiar Figures (MFF) procedure. In each, there is one standard and a number of comparison figures. One of the comparison figures is exactly identical to the standard, while the others differ from it in relatively small details. The subject must point to the picture which is identical to the standard. Matching Figures 1 involves one sample and eleven test trials. Five comparison figures appear on each trial. Stimuli are drawn from Scale 1 of Cattell's Culture Free Test of *g*. (Cattell, 1950) Matching Figures 2 uses material drawn from Kagan's MFF test. One sample and nine test trials are run. On each trial there are three comparison figures.

In administering the PPVT, the standard instructions for the vocabulary test were employed. It was assumed that reflection--impulsivity is measured here in a relatively unstressful context. Likewise, the Dots procedure was meant to be unstressful. Whatever response the child made was accepted as good, and he received praise and encouragement throughout.

The Matching Figures tests were designed to provide measures of reflection--impulsivity in an evaluative context. Here, also, the child gave only one response for each item. If he was correct, a marble was placed in a cup before him, and he was praised; if the choice was wrong, he received no marble and was told he had made an error. Before the test, the child had been shown a prize which could be won by good performance in the task--by being right often enough and thereby winning enough marbles. All children received their prize; but it was expected

that these procedures would involve considerably more stress than the others employed.

To avoid contamination of other procedures, the two reflection--impulsivity measures administered in an evaluative context were the final two given--Matching Figures 2 occupied the eighth session; Matching Figures 1, the ninth. The Dots test was the sixth given, and fell between the two forms of the PPVT.

Questionnaire

In the hope of obtaining information relevant to performance on the creativity measures, a questionnaire was sent to parents of all children tested. Several kinds of questions were included. One set called for descriptive information--parents' education and occupation, number of children, number of years the child had been in kindergarten or nursery school. A second set asked for parents' rating of typical behaviors--how aggressive, independent, obedient, energetic, happy, etc., they considered their child to be in relation to other children of his age. His play preferences were also the subject of several questions--what activities are preferred; does he prefer contact with other children or with adults; does he lead or follow; to what extent do other children seek him as a playmate. Several questions concerned discipline used by the parents--how often is physical punishment used; to what extent do both parents discipline the child; which of seven ways of disciplining the child were the parents most likely to employ. The latter

question was a modified form of one described by Getzels and Jackson (1962). Finally, the parents were asked about their goals for the child--both in the short run, and in terms of traits most wanted for him to develop.

The questionnaire was distributed to parents by mail in one case and by the school in the other three. It was accompanied by a letter describing its purpose, and by a ten-page mimeographed summary of the background, aims, methods, and potential values of the research being conducted. In two of the schools, this distribution came shortly after a parents' meeting at which the author discussed his research. In every case, the parents' meeting and the explanations were delayed until all of the creativity assessment and most of the other measures had been completed.

None of the questionnaire items showed relations to creativity which were both significant and consistent across sex or across the set of creativity measures. Hence, no further reference to questionnaire data will be made.

Scoring of Creativity Instruments

To facilitate the discussion of results, details of the creativity scoring will be presented here. Wallach and Kogan noted that "...the game-like instructional context did not eventuate in a flouting of the task constraints inherent in each experimental item. Bizarre or inappropriate responses--those judged inadmissible as possible answers to a given question--were extremely rare." (Wallach and Kogan, 1965, p. 37)

Thus they were able to obtain their total and total unique measures simply by counting responses. This procedure was not possible in the present investigation. A sufficiently large number of responses appeared to be unrelated to the stimulus item to require the use of judges in determining response admissibility.

Data were prepared for judging by printing each response on a separate IBM data processing card. Each card included a code identifying subject, procedure, and item. E totaled frequencies for the following classes of responses but did not prepare IBM cards for them:

Repetitions. This class included exact repetitions of responses within an item. Additionally, if both an item and a superordinate for that item were given, only the more specific response was scored, and the superordinate was counted as a repetition. An example would be ball and baseball; the former would be scored as a repetition. However, if a pair could but commonly would not have the superordinate--subordinate relationship, both responses were scored; for example, wheel and steering wheel.

Incomprehensible. This class consisted of all responses not understood by E because of unclear or incomplete expression or recording.

Bizarre. Those few responses which were grammatically or logically inadmissible were excluded as bizarre. Examples include "Man...he walks" as a response in naming round things; and "green ball" as a response for things that are red.

For all responses not excluded on the above grounds, it was decided to accept as appropriate any response given by two or more Ss in the sample. For this scoring, variations in number and in non-critical adjectives were not considered to differentiate responses. For example, the following were considered to be the "same" response: "hang a coat" and "hang coats," in uses for a coat hanger; "little circle" and "circle," in naming round things; "red leaf" and "leaf," in naming red things.

The cards bearing responses that were unique in the sample were presented to three independent judges. Each judge worked with all responses to one item before proceeding to the next. Each received the procedures in a different order, after the previous judge's decisions had been recorded and the cards for each item shuffled.

The judges were familiarized with the experiment and its purposes. They were asked to sort responses into three categories: "Those that you think are bizarre or inappropriate;" "Those that you think are appropriate or reasonable;" and "Those that you just can't make a decision on." They were instructed to keep the latter category to a minimum, and to give the child the benefit of doubt when expression was incomplete or relevance was marginal. No effort was made to give the judges an elaborate (and arbitrary) set of rules for handling difficult cases; rather, they were to categorize on the basis of whether they could see a relationship between task and response.

CHAPTER III

RESULTS

The presentation of results will be given in three major sections. In the first, creativity scores and creativity--intelligence relationships will be considered. The second section will treat results for the reflection--impulsivity dimension. Finally, the analyses searching for interrelations between these two domains will be presented.

Creativity

Scoring. The procedure used in scoring has been described in the preceding chapter. To simplify analysis, the few judgments in the "unable to decide" category were assigned to the "bizarre" group.

All three pairs of judges showed a satisfactory percent agreement for each of the fifteen items making up three creativity procedures. These percents ranged from 65.5 to 96.4. For the three pairs of judges, median percents agreement were 80.3; 76.2; and 92.0. There was no indication of different degrees of agreement for the different creativity measures. Finally,

the percent of responses for which all three judges agreed on the assignment ranged from 63.2 to 92.4, with a median of 74.6. Where the judges failed to concur, the response was assigned to the category agreed upon by two of the three.

The percents of total responses falling into various categories are presented in Table IV.[1] Of these totals, the overall percent of responses discarded ranged from 13.78 for the Pictures test to 22.16 for Uses. For Instances and Uses, the largest single basis for response rejection was repetition; for Pictures, most of the responses discarded were those judged "bizarre" by the raters. Considering only admissible responses, the percent of responses that were unique in the sample ranged from 21.08 for Instances to 39.15 for Uses.

Reliability of creativity scores. Item-sum correlations for unique and for nonunique responses indicate that the various creativity items contribute uniformly to response totals.[2] Median reliability coefficients appear in Table V for all subjects and separately for males and females. No systematic differences appear when coefficients are examined separately by sex. For all subjects, these correlations range from .42 to .92 for unique responses, and from .40 to .88 for nonunique responses.

---

[1] Ninety subjects completed the creativity procedures and are included in the analyses reported here. Three males at School D did not complete the entire test battery. Their data are excluded from all subsequent analyses.

[2] "Unique" responses are those which were unique but judged admissible. Because the "unique" and "total" indexes are not independent, most analyses replace the latter with the total corrected by subtracting unique responses from it. These totals are referred to as "nonunique" responses.

Table IV

Categorization of Responses in Creativity Tasks[a]

	Procedure		
Category	Pictures	Instances	Uses
Rejected in Coding	3.40	12.27	19.56
Repetitions	1.58	8.11	10.03
Incomprehensible	1.09	0.86	1.78
Bizarre	0.73	3.30	7.75
Rejected by Judges	10.38	4.51	2.60
Appropriate--Total	86.22	83.22	77.84
Unique	25.02	17.55	30.48
Nonunique	61.02	65.67	47.36

[a] Numbers in table are percents of total responses for the procedure.

Table V

Median Item-Sum Reliability Coefficients

		Procedure		
Measure	N	Pictures	Instances	Uses
Unique Responses				
All Subjects	87	.54	.89	.89
Males	41	.48	.85	.92
Females	46	.60	.91	.74
Nonunique Responses				
All Subjects	87	.48	.77	.65
Males	41	.52	.77	.74
Females	46	.54	.75	.56

Those for the Pictures test are consistently lower than those for the other two.

After completion of the test battery, the Pictures test was readministered to subjects at School A. The author was the sole judge of response admissibility, and responses were not added to the main Pictures data. For eight males and fifteen females, test--restest correlation for total acceptable responses was .79. No test--retest data were obtained for the other procedures.

Creativity totals. Means and standard deviations of creativity scores are presented in Table VI. For each sex, means are given for unique responses, nonunique responses, and those rejected. One-way analysis of variance was used to test for sex differences in these scores as well as in total responses, total admissible responses, number of repetitions, and number of bizarre or incomprehensible responses.[1] In twenty-one analyses there appeared only one *F* significant at the five percent level; hence, no sex difference is indicated.

Several additional breakdowns of the data were tested by analysis of variance. Means for younger subjects (below six years) were compared with those for subjects older. For neither sex did any difference approach significance at the five percent level.

Possible differences among samples drawn from the four schools were examined. As above, analyses were performed

---

[1] Response distributions were positively skewed. However, no difference in pattern or magnitude of relations appeared when the major creativity analyses were repeated using the log (X+1) transformation, which eliminated the skew in most cases. Therefore, analyses based on raw scores are reported.

Table VI

Mean Creativity Scores

		Males (N=41)		Females (N=46)	
Procedure	Measure	Mean	S.D.	Mean	S.D.
Pictures	Unique	4.78	3.06	4.33	3.72
	Nonunique	11.12	3.02	11.46	3.61
	Reject	2.54	3.08	2.57	7.44
Instances	Unique	4.10	4.33	3.57	6.79
	Nonunique	14.73	8.85	14.30	8.22
	Reject	3.54	4.44	3.89	9.16
Uses	Unique	6.15	8.89	4.26	3.62
	Nonunique	9.10	3.67	7.57	2.47
	Reject	4.17	6.28	3.70	7.60

separately by sex and with each of the seven scores listed. None of the forty-two Fs reached significance at the .05 level.

Finally, analyses based on Hollingshead's index of socio-economic status were performed, using the fifty-eight subjects on whom scores were available. Data for males and females were combined, and three status groups were formed: those above, in, and below Class III. No significant differences appeared in these comparisons.

Creativity and intelligence. Correlations among creativity scores and those between the creativity and intelligence domains are reported in this section. In these analyses, each creativity measure is represented by two statistically independent scores--total unique responses, and total that are nonunique. Convergent intelligence scores are MA and IQ obtained by averaging results from the two forms of the PPVT. Test--retest correlation for the latter were .61 for males and .67 for females, for the IQ index; for MA they were .60 and .65. Significance levels for correlations were obtained by Fisher's r-to-z method.

Product moment correlations for all 87 subjects are shown in Table VII. Eight of fifteen correlations within creativity are significant, all at the .01 level, including five significant at .001. However, only one of eight correlations between scores on the Pictures test and those on Instances or Uses is significant. It appears that the Instances and Uses tests hold much of their variance in common; the six correlations within these procedures have a mean of .44 ($p < .001$), which is comparable to values presented by Wallach and Kogan (1965). The

Table VII

Creativity and Intelligence Correlations for all Subjects[a]

Procedure	Pictures--Nonunique	Instances--Unique	Instances--Nonunique	Uses--Unique	Uses--Nonunique	MA	IQ
Pictures--Unique	36***[b]	16	17	08	-03	19	11
Pictures--Nonunique		16	29**	13	-06	21	13
Instances--Unique			73***	43***	28**	-09	-14
Instances--Nonunique				51***	42***	-04	-13
Uses--Unique					29***	-01	-08
Uses--Nonunique						-02	-18

[a] Decimal points are omitted. N=87.

[b] * $p < .05$
** $p < .01$
*** $p < .001$

Pictures test, however, bears little relation to these.

None of the measures employed show a substantial relation to the intelligence measures. Wallach and Kogan have suggested that MA, rather than IQ, should be compared with creativity scores, since like the latter it represents a score uncorrected for age. Use of a corrected score for one variable and an uncorrected one for the other may artificially curtail relationships between the two domains. It will be observed in Table VII that correlations between creativity and IQ are higher (more positive or less negative) when MA rather than IQ is employed, following from the nonsignificant positive correlations (mean is .16) between age and creativity scores. In each case, however, it is clear that correlations between the Instances and Uses tests are not derived from dependence of these scores on intelligence. Low negative correlations are found between MA and scores on these two tests, while those between Pictures and MA are positive and nonsignificant.

The correlation matrix for 41 males is presented in Table VIII. Creativity intercorrelations match closely those obtained for the sample as a whole. Correlations involving only the Instances and Uses tests have a mean of .57 ($p < .001$), while those between these tests and Pictures show a mean of .06. All of the former are significant, at the .001 level in five of six cases; while none of the latter approach significance.

The significant positive correlations do not result from correlations with IQ. Scores on the Instances and Uses tests have consistent low negative correlations with MA. On the other

Table VIII

Creativity and Intelligence Correlations for Males[a]

Measure	Pictures--Nonunique	Instances--Unique	Instances--Nonunique	Uses--Unique	Uses--Nonunique	MA	IQ
Pictures--Unique	41**[b]	04	17	05	09	52***	41**
Pictures--Nonunique		-05	16	04	01	12	08
Instances--Unique			77***	55***	58***	-30	-38*
Instances--Nonunique				58***	59***	-19	-30
Uses--Unique					33*	-19	-27
Uses--Nonunique						-17	-38*

[a] Decimal points are omitted. N=41.

[b] * $p < .05$
** $p < .01$
*** $p < .001$

hand, a significant positive correlation appears between intelligence scores and responses on the Pictures test.

Correlations for the 46 females are presented in Table IX. Those involving only the Instances and Uses tests average .35 ($p < .05$), while those between Pictures and the other two average .16. Correlations between creativity and MA average .12. These are generally positive, and two are significant.

Females, like males, show greater agreement among measures derived from the Instances and Uses tests than can be accounted for by relations with intelligence. Three of the six creativity intercorrelations here are significant, each at .001; while only one of four creativity--MA correlations is significant.

Mean correlations are weaker for females than for males, but in each sex there is evidence that two of the creativity measures have substantial variance in common, and that this variance is not shared with the measure of convergent intelligence. However, there is no evidence that the Pictures test is measuring the same dimension of individual difference as that measured by Instances and Uses. No significant correlations between Pictures and the others are found for males, and only two for females. For males, also, the Pictures test has positive correlations with MA, one of them significant; while Instances and Uses have consistent low negative correlations with MA.

The failure to replicate the correlational results of Wallach and Kogan may indicate that a unitary creativity dimension is not to be found in children of kindergarten age.

Table IX

Creativity and Intelligence Correlations for Females[a]

Measure	Pictures--Nonunique	Instances--Unique	Instances--Nonunique	Uses--Unique	Uses--Nonunique	MA	IQ
Pictures--Unique	34*[b]	22	18	13	-20	-07	-12
Pictures--Nonunique		27	41**	37*	-12	30*	19
Instances--Unique			74***	50***	07	01	-03
Instances--Nonunique				48***	21	11	01
Uses--Unique					12	29*	20
Uses--Nonunique						06	-07

[a] Decimal points are omitted. N=46.

[b] * $p < .05$
** $p < .01$
*** $p < .001$

On the other hand, it is possible that the results for the Pictures test are a function of the test rather than of the subjects. Alternative interpretations that are consistent with the data will be considered in the Discussion section. In the following pages, additional analyses of the creativity data are presented.

Creativity and intelligence for brighter children. One possibility to be considered is that the failure of Pictures to correlate with the other creativity procedures results from the relatively low intelligence level for the sample. Torrance, Roe, and MacKinnon, among others, have suggested that a cutoff point important for relations between creativity and other variables will be found at around an IQ of 120. Wallach and Kogan did not present an average IQ for their subjects, but extrapolation from the mean WISC scores and mean ages they give indicates that this average may be as high as 120. In the present sample, mean IQ was about 103; eleven males and twenty-one females, or 37% of the subjects, had IQs of less than 100.

An argument, admittedly *post hoc*, can be made that the Pictures test imposes more stringent requirements for acceptable responses than the two other creativity procedures, so that this test in particular may be affected by the low intelligence level for the sample. In the first place, it appears that greater abstractive ability is required for responding to Pictures than to the other procedures. For each picture, there may be one or a few responses that can be derived straightforwardly from the child's experience; e.g., "umbrella" for pattern 3; "sun" for pattern 4 (see Figure 1, p. 85a). High scores,

however, demand that the child go beyond these to responses for which the stimuli are less obviously representations. Thus, he must produce responses whose essential features only are suggested by the stimulus. On the remaining tasks, high scores can be obtained without such abstraction. The categories for Instances are concrete and such that their critical features appear, to the adult at least, to be perceptual "givens." For the Uses test, also, responses can be drawn directly from experience with the objects in question and with similar objects.

It is possible, also, that responding to Pictures requires a different and more sophisticated form of search through memory than is needed for the other tasks. For Instances, perceptually present objects meeting the criteria of "roundness" or "redness" could be, and often are, given; perception then provides a set of possible responses, and the task can be made one of discriminating among these, rather than of generating a set of response possibilities. Similarly, for Uses, ordinary experience with the objects in question could provide a pool of readily available responses. It is difficult to imagine so ready a reservoir of potential responses for most of the Pictures items--no clue is given as to what realm of experience will provide appropriate items, and the essential features of the stimulus do not correspond to a simple attribute ordinarily used to classify objects. Perhaps, then, the child must search more widely through his experience to find a basis for Pictures responding.

This argument must rest on its face validity, since we know little or nothing about the organization of memory or about

the way in which a "creative" response is related to what is present in storage. Less speculatively, it may be argued that if responding to Pictures is more difficult than responding to the other procedures, subjects should produce fewer acceptable responses per stimulus item on Pictures than on the others. This is the case; across all subjects the mean number of responses per item is 2.0 for Pictures, 3.4 for Uses, and 6.1 for Instances.

It is possible, then, that greater coherence of creativity results will be found when analysis is restricted to subjects of above-average intelligence. This possibility was examined by repeating the correlational analyses on groups from which subjects below 100 in IQ were excluded. It would be most desirable to perform a series of such analyses, with a step-by-step increase in the cutoff level for IQ, and to proceed up to a baseline of 120 or above; in view of the present sample size, however, only this one analysis seems justified. As seen in Table X, this procedure leaves a reasonable sample size (thirty males and twenty-five females), while yielding groups with mean IQ of about 111. These groups should at least approach those of Wallach and Kogan in intelligence level. It may be noted also that there is a substantial reduction in IQ variance for the restricted groups, but that enough variation should remain to permit relationships between creativity and intelligence measures to appear.

Correlations are presented in Table XI for males and in Table XII for females. For males with IQs of 100 or greater, nine of fifteen correlations among creativity measures are significant, including five at .01 or better. Mean correlation is

Table X

Intelligence Means and Standard Deviations

Group	N	Intelligence Measure Mental Age Mean	 S.D.	 IQ Mean	 S.D.
No IQ Restriction					
All Subjects	87	73.39	13.22	102.87	13.66
Males	41	76.50	13.31	105.38	13.12
Females	46	70.62	12.65	100.63	13.89
Excluding Subjects with IQ < 100					
All Subjects	55	80.69	9.58	111.08	8.25
Males	30	81.70	11.23	111.33	9.20
Females	25	79.48	7.15	110.78	7.11

Table XI

Creativity and Intelligence Correlations for Males Excluding those with IQ less than 100[a]

Measure	Pictures--Nonunique	Instances--Unique	Instances--Nonunique	Uses--Unique	Uses--Nonunique	MA	IQ
Pictures--Unique	52**[b]	19	41*	09	45*	51**	34
Pictures--Nonunique		25	59***	14	35	17	09
Instances--Unique			68***	53**	37*	-22	-23
Instances--Nonunique				50**	42*	-01	-11
Uses--Unique					28	-20	-32
Uses--Nonunique						26	05

[a] Decimal points are omitted. N=30.

[b] * $p < .05$
** $p < .01$
*** $p < .001$

Table XII

Creativity and Intelligence Correlations for Females Excluding those with IQ less than 100[a]

Measure	Pictures--Nonunique	Instances--Unique	Instances--Nonunique	Uses--Unique	Uses--Nonunique	MA	IQ
Pictures--Unique	-10	21	26	10	-08	15	-07
Pictures--Nonunique		23	26	37	-18	16	10
Instances--Unique			86***[b]	61***	13	-08	-26
Instances--Nonunique				67***	32	-05	-21
Uses--Unique					25	17	01
Uses--Nonunique						04	-06

[a] Decimal points are omitted. N=25.

[b]
* $p < .05$
** $p < .01$
*** $p < .001$

.38 ($p < .05$). Correlations involving only Instances and Uses average .46 ($p < .05$). Thus, within the restricted sample, these procedures continue to hold much of their variance in common.

Correlations between measures from the Pictures test and those from Instances and Uses average .31. With the reduced sample size, this mean correlation reaches only the .10 probability level. However, three correlations are significant at .05, one of them reaching the .001 level; and all are positive.

Despite the exclusion of low intelligence subjects, Pictures totals are in part a function of intelligence level. Unique responses on Pictures correlate with MA at .51 ($p < .01$). MA has low correlations with measures derived from Instances and Uses, with three of these negative. Hence, the significant correlations between Pictures and the other measures are not a function of relations with intelligence.

Results for males are compatible with the interpretation that a unitary dimension is present, but that the Pictures test, becsuse of its greater difficulty, is substantially related to this dimension only when relatively low intelligence subjects are excluded. Even then, with five and six year olds, Pictures performance is determined jointly by intelligence and "creativity." The Instances and Uses procedures, on the other hand, are strongly intercorrelated and distinct from intelligence whether or not below--average children are included.

For females, however, elimination of low intelligence subjects does not lead to unity within the creativity domain. Mean

correlation among creativity measures is .26, with only three correlations signficant. Correlations among the Instances and Uses measures average .47 ($p < .02$), with each of the six coefficients slightly higher for the restricted group than was the corresponding coefficient for all females. Thus, for these two procedures, creativity measures show somewhat greater coherence when low IQ subjects are removed. Correlations of the Pictures test with Instances and Uses, on the other hand, range from .18 to .37, with a mean of .15. They are thus no different from correlations obtained across all females; and with the reduced sample size, none of the coefficients is significant. Finally, it may be noted that creativity measures have a mean correlation of .07 with MA, and no correlations are significant. Hence, the positive intercorrelations for Instances and Uses are again not the result of relationships with intelligence.

Elimination of low intelligence subjects, then, does not produce a correlational pattern equivalent to that found by Wallach and Kogan. Pictures totals correlate with MA as well as with other creativity measures, in boys; they fail to correlate with Instances and Uses in girls.

Involvement in the task. These data leave viable the possibility that the creativity dimension is a meaningful one, but that the Pictures test is an inappropriate measure at this age. Some evidence for dimensional unity may be seen in measures of involvement in the creativity tasks. Wallach and Kogan have proposed that one aspect of creativity is a high degree of involvement, with unwillingness to relinquish a problem until all

possibilities have been exhausted. It may be that the children studied will show consistency in their degree of involvement from one creativity task to another, despite failures of consistency in response production. Such a finding would constitute evidence for the presence of varying degrees of a disposition related to the dimension studied by Wallach and Kogan, and would support the interpretation that failures of correlation for the Pictures task reflect characteristics of this task rather than the absence of the dimension.

Involvement in the task is inferred from time spent in responding. It will be recalled that the child continued in the task until he indicated, for each item, that he had no more responses to give. The child with greater involvement should continue attempting to respond for a longer period of time, whether or not he can fill this time with appropriate responses.

Time was recorded in ten-second intervals from the presentation of the item until the end of the last interval in which a response was given (regardless of whether the last response was judged to be admissible). There is thus some necessary dependence of time on responding; but this dependency is slight, since children can respond at greatly different rates.

Time in task varied greatly from subject to subject. When summed over all items in a procedure, it ranged from six to 111 ten second periods for the Instances test; from two to 112 for Uses; and from five to 92 for Pictures. Medians ranged from 16 for Uses to 21 for Pictures, and were highly similar across sex. The positive skew in these distributions was largely eliminated

by the log(X+1) transformation, used on time data for all correlations to be reported.

Correlations among the time measures are presented in Table XIII, separately by sex and with breakdowns into low and high IQ groups. Totals for all males and for all females yield significant correlations among all three procedures. Mean correlation is .55 in each sex (for males, $p < .01$; for females, $p < .001$). When high and low IQ subjects are considered separately, magnitudes of the correlation coefficients are comparable to those for subjects covering the whole range studied, although with reduced sample sizes three of twelve correlations fail to reach significant. Thus, time spent in responding to the task yields a measure showing greater consistency across tasks and IQ range than is shown by response totals on those tasks. The correlations for males with IQ below 100 are particularly striking. For these subjects correlations between totals on Pictures and those on the other creativity procedures were all negative, ranging from -.01 to -.41, and with a median of -.27. Nonetheless, time on the Pictures test has positive correlations on the order of .5 with time for Instances and Uses.

Time in task is not a function of intelligence level. For females, correlations between log time scores and MA range from .07 to .09. For males, negative correlations on the order of -.30 are found for Instances and Uses, while the correlation for Pictures is .13.

Correlations between time and response totals are given in Table XIV. The highest correlations are those between measures

Table XIII

Correlations among Time in Task for Creativity Procedures[a]

	Males			Females		
Procedures	All (N=41)	IQ < 100 (N=11)	IQ ≥ 100 (N=30)	All (N=46)	IQ < 100 (N=21)	IQ ≥ 100 (N=25)
Pictures x Instances	44**[b]	51	48**	51***	52*	53**
Pictures x Uses	36*	55	37*	47***	60**	37
Instances x Uses	86***	93***	83***	68***	74***	64***

[a] Decimal points are omitted.

[b] * $p < .05$
** $p < .01$
*** $p < .001$

Table XIV

Correlations between Time and Response Measures[a]

Response Totals	Time in Task					
	Males			Females		
	Pictures	Instances	Uses	Pictures	Instances	Uses
Pictures--Unique	60***[b]	14	08	71***	30	24
Pictures--Nonunique	51***	08	01	63***	46**	40**
Instances--Unique	46**	66***	63***	30*	76***	59***
Instances--Nonunique	50***	76***	73***	37*	82***	59***
Uses--Unique	26	50***	74***	34*	45**	62***
Uses--Nonunique	34*	49***	48**	-09	09	33*

[a] Decimal points are omitted. Time scores have been transformed by log (X+1). Correlations are for 41 males and for 46 females.

[b] * $p < .05$
** $p < .01$
*** $p < .001$

derived from the same procedure, but time for each procedure generally shows a substantial relation to number of responses produced on each other procedure. For males, the sole exception to this pattern is the failure of response totals on Pictures to correlate with time measures derived from Instances and Uses. The same failure appears for females, but only for unique responses to Pictures.

Time in task is thus highly consistent across all three tasks, and shows substantial correlation with response totals for two of them. These results suggest that a common dimension underlies performance on the creativity tasks, although this dimension is not clearly manifested in response totals on Pictures.

The persistence interpretation. An alternative interpretation suggested by the results presented above must be considered. Given the lack of control of time spent in responding to the creativity tasks, and the generally positive correlations between time and response totals, it might be argued that high response totals are simply a function of persistence in the task. It is clear that no complete separation between "persistence" and "creativity" as interpretations of high totals can be made, since, in Wallach and Kogan's formulation, the tendency to stay with the task until exhausting its response possibilities is an integral part of creative performance. However, it is reasonable to ask whether creativity totals represent "nothing but" the tendency to persist in the creativity task; or, indeed, the tendency to persist in any task.

Evidence that creativity totals do not result from a general tendency to persist comes from correlations with time spent in the drawing task. Here, time was measured in seconds from completion of the instructions until the subject indicated that his drawing was finished. It ranged from 26 seconds to twenty minutes, with a median of 118 seconds for boys and 142 seconds for girls.

For each sex rank order correlations were computed between drawing time and total fluency score on each of the creativity measures. These correlations ranged from -.11 to .13, with a mean of .01. There was, then, no general "persistence" appearing in both the creativity measures and the drawing procedure, although these situations are similar in their open-endedness and lack of time constraints.

There is also some evidence with regard to persistence specific to the creativity tasks. This is obtained for the Instances procedure, where the fluency score was recalculated deleting all responses occurring more than two minutes after the introduction of each item. The two minute time limit was chosen since it has been used frequently, particularly in Guilford's batteries. If fluency within this time period correlates with total fluency for Instances, persistence will not account for response totals.

Correlations were obtained separately by sex for those below an IQ of 100 and for those of 100 or better. Number of admissible responses within the first two minutes of responding had correlations with total admissible responses ranging from

.80 for low IQ females to .89 for high IQ males. Fluency with time restricted also showed correlates with total fluency on the other two creativity procedures that were highly similar to those obtained without the time restriction. The mean difference in correlations, when restricted fluency is substituted for total fluency, was .05 for the male groups and .02 for females.

There is clearly a need for better evidence here. Such evidence as there is suggests, though it does not compel, the conclusion that neither general nor task-specific persistence is sufficient to account for number of responses produced in the creativity procedures. It appears, then, that time spent in responding on the creativity tasks is a correlate rather than a determinant of high fluency totals on the Instances and Uses procedures.

Response quality. Several additional analyses of scores derived from creativity measurement were conducted. One of these concerned the degree of consistency shown by subjects in the number of unacceptable responses given, and the relationship of these totals to creativity and intelligence measures. For each sex, number of bizarre and incomprehensible responses and number of repetitions had consistently positive correlations within and between tasks. Total responses rejected intercorrelated significantly across all three tasks for females, and had two of three correlations significant for males, all at the .01 level or better. Each of these totals was negatively correlated with MA, with coefficients ranging from -.27 to -.61

for males, and from -.23 to -.30 for females.

Number of responses discarded generally showed positive correlations with number of acceptable responses given. These correlations are given in Table XV for total acceptable responses and for total responses discarded; the correlations were similar when unique and nonunique responses were considered separately, and when repetitions and other unacceptable responses were separated. For males, responses discarded on each test yield positive correlations with admissible responses given for Instances and Uses, while showing no such relations to number of acceptable responses given on the Pictures test. For females, however, all correlations are positive, and admissible responses on Pictures have consistently high correlations with responses rejected.

Positive correlations between number of acceptable responses and number of unacceptable ones could indicate that children who achieve high scores on creativity instruments do so because they give many marginally acceptable responses; that is, that quality and productivity will be negatively related. To test this possibility, an examination was made of responses to the Uses test, whose items are most susceptible to objective scoring for quality.

One index to low response quality is the percent of responses given which represent use of the object in highly conventional ways. Thus, there are a number of uses for a knife that involve cutting, slicing, or spreading food. In the scoring employed, "cut meat" and "cut butter" were counted as

Table XV

Correlations between Responses Accepted and Those Discarded[a]

	Males			Females		
	Pictures	Instances	Uses	Pictures	Instances	Uses
Pictures--Total	01	-21	-23	66***[b]	61***	39**
Instances--Total	55***	50***	47**	20	63***	57***
Uses--Total	16	28	73***	13	12	16

---

[a] Decimal points are omitted.

[b] * $p < .05$
** $p < .01$
*** $p < .001$

distinct responses for the total fluency score, while "cut a banana," which was given by only one child, received credit as a unique response as well. If high scores are derived from greater listing of uses limited to such conventional categories, then they are an indication of production we may not wish to call creative.

Responses falling into the most frequent category employed were summed across the four subparts of the Uses test. These categories were, for newspaper, uses related to information exchange (reading, looking at pictures, use as TV guide, advertising things for sale); for knife, uses related to food; for cup, things related to drinking or containing liquids; and for coat-hanger, uses concerned with hanging clothing or other objects. Across all subjects, these four categories accounted for 41.8% of all acceptable responses, including 60.7% of nonunique responses and 16.7% of those that were unique. No differences appeared in these percents between males and females, or between more and less intelligent subjects. Of primary interest here, the fluency total on Uses was uncorrelated with percent of responses falling into these four categories. The correlation was -.07 for males, and -.09 for females.

Thus, high scores do not represent a tendency to give more responses belonging to highly common categories. All subjects give many such responses, but totals evidently depend on the number of responses given exclusive of these categories. Further, greater fluency is associated with a greater proportion of responses which are both unique and acceptable. For the

various procedures, correlations between total admissible responses and percent unique responses range from .23 to .62. Mean correlation is .48 for males and .45 for females ($p < .01$ in each case). It appears, then, that high response totals are not the result of giving many marginally acceptable responses.

Category scoring. A more intensive examination was made of the relation between number of responses and number of categories employed on the Uses procedure. It was found that most responses given could be placed in one of a few categories. For "uses for a cup," for example, six categories were sufficient--uses related to drinking; use as container (other than for liquids); use as a toy; use as a weapon; use for or in decoration; and household uses. This latter category may be more comprehensive than the others; but, since few subjects gave many responses falling into this category, no further differentiation appeared to be necessary.

Table XVI presents correlations based on the category scoring performed by the author; scores derived by an independent judge following his rules correlated with his at .89 for males and .86 for females. For males, number of categories employed in the Uses procedure is substantially correlated with totals on Instances and Uses for high and low IQ subjects, while correlating with the Pictures total only for subjects with IQ greater than 100. Correlations with MA cannot account for these relations. For females, category scores are more test-specific, and are related to MA as strongly as to fluency totals on the other procedures. This pattern of results is compatible with

Table XVI

Category Scores on Uses

Correlations with Fluency and Mental Age Measures[a]

Group	N	Measures			
		Pictures	Instances	Uses	MA
All Males	41	32*[b]	49***	45**	07
M,IQ $<$ 100	11	-04	32	50	-02
M,IQ $\geq$ 100	30	42*	53**	45*	18
All Females	46	-07	23	59***	52***
F,IQ $<$ 100	21	-11	29	45*	25
F,IQ $\geq$ 100	25	29	19	67***	36

[a] Decimal points are omitted.

[b] * $p < .05$
** $p < .01$
*** $p < .001$

that seen in the fluency intercorrelations. For males with IQ over 100, the creativity measures hold much of their variance in common, regardless of the scoring employed--fluency, flexibility (number of categories), or uniqueness. For females, creativity measures are less strongly related to one another than is the case for males.

Summary of creativity results. Interjudge agreement on admissibility of unique responses, and item-sum reliability of uniqueness and fluency totals, were seen to be adequately high for all three creativity procedures employed. Response totals bore no significant relationship to age, sex, school, or socioeconomic status within the sample.

For the sample as a whole, there was no evidence that the creativity measures were tapping a unitary dimension of individual difference. In each sex, measures derived from the Instances and Uses procedures were strongly related to one another, and were much less strongly related to convergent intelligence; but the third creativity measure was virtually independent of these. However, there was some support for the interpretation that the failures of correlation for the Pictures measures reflected inappropriateness of the procedure rather than lack of the dimension. For males, response totals on Pictures were seen to be a joint function of "creativity" and intelligence when low IQ subjects were excluded from the analysis. For both sexes, time spent in responding on the creativity procedures, interpreted as a measure of task involvement, was highly consistent across all tasks. Both low and high IQ subjects showed this consistency,

and the time measures were strongly related to fluency totals on the Instances and Uses procedures.

The time and response correlations do not imply that creativity totals result from either a general or a task-specific tendency toward "persistence." Time spent in drawing, another free and open-ended procedure, was unrelated to response totals; and response fluency within the first two minutes of responding on the Instances procedure had the same correlates as the fluency measure with no cutoff time imposed.

Response quality was examined in several additional analyses of data from the Uses procedure. All subjects were seen to give many responses representing highly conventional uses of the objects employed; but greater use of conventional categories did not account for higher fluency totals. Rather, greater fluency was associated with giving proportionately more responses which were both unique and appropriate. Finally, for males, number of categories employed had correlates similar to those of the fluency totals for the Uses procedure. For females, this generality did not obtain, but greater fluency on Uses was associated with the use of a greater number of categories. In both sexes, then, there is some evidence that greater fluency is associated with greater diversity and uniqueness of response, rather than resulting from extensive listing of highly conventional or marginally relevant responses.

Reflection--Impulsivity

In this section the reflection--impulsivity results are presented. Discussion of relations of impulsivity to the creativity findings is deferred to the section following.

It will be recalled that two different administration contexts were employed in measuring reflection--impulsivity. Matching Figures 1 and 2, the procedures most similar to those used by Kagan, were given under "evaluative" conditions--the subject was informed whether his response was correct, and a prize was apparently contingent on a high proportion of correct responses. On the other hand, the usual "permissive" atmosphere obtained for the administration of Forms A and B of the PPVT, and for the Dots procedure. Here, praise and encouragement were liberal; there was no feedback as to errors, and no contingency of prize on performance. It was hypothesized that some subjects would perform differently under these contrasting conditions, while others would be unaffected. Nonetheless, if reflection--impulsivity represents a pervasive dimension of individual difference, latencies should be intercorrelated across all measures; though the greatest consistency may emerge for tests given in the same context.

Response time distributions from the five tests of reflection--impulsivity were converted to normalized _T_ scores with mean of 50 and standard deviation of 10. Normalization and all correlational analyses were performed separately for males and females. Median tests on the raw data, however, showed no sex differences for the latency scores, and only one

significant difference on error scores: females made significantly fewer errors than males on the Dots test ($p < .01$).

Correlations among latencies are presented in Table XVII for males and Table XVIII for females. Eight of ten intercorrelations are significant for males, with a mean intercorrelation of .38 ($p < .05$). For females, seven of ten are significant, and the mean is .33 ($p < .05$). These correlations are of the same order of magnitude as those presented by Kagan, despite the context differences and the employment here of three untested measures which are phenotypically quite different from those used earlier.

It is apparent that impulsivity intercorrelations are not a function of relationships with intelligence. For males, mean correlation with MA is .13 and no correlations are significant; mean correlation is .20 for females, with two correlations significant.

For each sex the highest correlation is that between the two measures given in an evaluative context. Matching Figures 1 and 2 are correlated at .64 for males and .57 for females ($p < .001$ for each sex.). Since these tests are also those that are most similar to the ones previously used, however, the higher correlations may simply be a function of greater reliability. That this is the case may be indicated by the correlations across contexts. For males, mean correlation among the three (new) measures administered in a permissive context is .36, while that between contexts is .35. For females, the means are .38 and .27. Measures administered in a permissive setting,

Table XVII

Reflection--Impulsivity Intercorrelations and Correlations with Intelligence for Males[a]

Measure	PPVT-A	PPVT-B	Matching Figures 2	Matching Figures 1	MA	IQ
Dots	48**[b]	23	34*	42**	23	21
PPVT-A		35*	37*	24	21	22
PPVT-B			43**	31*	03	01
Matching Figures 2				64***	09	23
Matching Figures 1					08	15

[a] Decimal points are omitted. N=41.

[b] * $p < .05$
** $p < .01$
*** $p < .001$

Table XVIII

Reflection--Impulsivity Intercorrelations and Correlations with Intelligence for Females[a]

Measure	PPVT-A	PPVT-B	Matching Figures 2	Matching Figures 1	MA	IQ
Lots	29*[b]	31*	40*	14	35*	29*
PPVT-A		53***	30*	18	02	-01
PPVT-B			36*	20	06	-05
Matching Figures 2				57***	23	23
Matching Figures 1					35*	39**

[a] Decimal points are omitted. N=46.

[b] * $p < .05$
** $p < .01$
*** $p < .001$

then, correlate with those given under evaluative conditions to about the same degree that they correlate with one another. Moreover, if the subject's normalized latencies are averaged across tasks within a context, greater reliability of measurement leads to across-context correlations of .52 for males ($p < .001$) and .38 for females ($p < .01$).

For neither sex does latency of response increase with age, when mean normalized latencies for tests given in the same context are compared by median test. Impulsivity intercorrelations, however, tend to be higher for older than for younger children. Mean correlation is .35 for younger males (less than 72 months of age), and .43 for older males. For females it is .25 for subjects under 70 months of age, and .40 for those above 70. None of these differences is significant.

Latencies do not differ when subjects above and those below the median IQ are compared by median test. For males, mean impulsivity intercorrelation is .27 for subjects below the median IQ of 105, and .42 for those above. For females, mean correlation is .39 for those below 101, the median IQ, and is .30 for those above. These differences also are not significant.

Kagan has reported negative correlations between number of errors on impulsivity measures and both latencies and IQ scores. Error scores were available on three of the five measures employed--Dots, and the two Matching Figures tasks. For each sex, these errors showed weak positive intercorrelations, with two of three correlations significant for males and none significant for females. Of the fifteen latency-error correlations, all

were negative for males, with five significant; for females, twelve of fifteen were negative, two of them significantly so. The best indication of latency-error relationship may come from the correlation between mean latency and mean error taken across all available measures. This correlation is -.44 for males ($p < .01$) and -.35 for females ($p < .05$). Thus, when reliability of measurement is increased by averaging scores from the individual tests, a substantial negative correlation between time and error scores is obtained.

Errors had negative correlations with mental age scores. These correlations are low for males, averaging -.15 and with none significant. They are stronger for females, averaging -.33 and with two significant. The average error score had a significant negative correlation with mental age in each sex-- -.32 for males ($p < .05$), and -.46 for females ($p < .01$).

These correlations provide additional evidence that the reflection--impulsivity dimension can be measured in kindergarten children, and that it contributes to task performance in the same way for these children as for older ones. Contrary to prediction, however, there is no indication that the administration context affected performance. Response latencies in a supposedly "permissive" context correlated with those from an evaluative context to the same extent that they correlated with one another.

It is possible that the tasks were incompatible with the effort to establish a nonevaluative setting. In both Dots and the PPVT, there is clearly a right answer for each item; the

failure to inform subjects of their errors may not have been as important as their awareness that there were correct and incorrect response alternatives available to them. Informal observation of subjects suggested that this was not the case. Subjects appeared to see the Matching Figures tasks as of much greater difficulty than the others on which impulsivity was measured; they began to make spontaneous comments, for example, that "This game is hard." Additionally, impulsivity measures under nonevaluative conditions were obtained immediately following the creativity assessment, in which the effort to establish a permissive atmosphere appeared to be highly successful. The evidence presented seems most compatible, then, with the conclusion that the child's reflectiveness or impulsivity is not affected by differences between the permissive and evaluative contexts of test administration.

Creativity and Impulsivity

No overall relationships were found between impulsivity scores and either time or response measures on the creativity procedures. No such relationships had been anticipated; rather, predictions were made for several subgroups derived by joint consideration of creativity and intelligence scores. It was intended to follow Wallach and Kogan's procedure of splitting the creativity and intelligence distributions at their respective medians to form four groups, highs and lows on one variable by highs and lows on the other. However, the failure of creativity

measures to intercorrelate over the whole range of IQ made the use of a composite creativity index unacceptable. It was also undesirable to discard subjects with IQs under 100 and then perform median splits on those remaining, although this procedure would have been legitimate for males. Aside from effecting a substantial reduction in the number of subjects available for analysis, this procedure would lead to a strange conception of a "low intelligence" group: low intelligence subjects would be those of a middle level of IQ for the sample.

Several different analyses were undertaken in an attempt to circumvent this difficulty. In one of these, analyses were performed separately for each creativity procedure. Within each sex the uniqueness and total fluency scores were summed, using weights inversely related to the standard deviations, to give a weighted creativity score for each of the three procedures. Median splits on these scores provided high and low creativity groups. Subjects were also grouped by three levels of intelligence--those with IQ below 100, those between 100 and 110, and those above 110. Three by two analyses of variance of the various reflection--impulsivity scores were performed, grouping in turn on the basis of each weighted creativity score, and using the method of unweighted means (Snedecor, 1946).

The results of this analysis were almost uniformly negative. The only consistent significant relationships with latencies appeared for females, where higher IQ was associated with greater reflectiveness on the Matching Figures procedures. None of the creativity main effects were significant, and only one creativity-

intelligence interaction reached significance for each sex. Average error scores tended to be higher for low IQ females in each analysis, though never significantly so; while they were higher for low creative males in each case, at the .05 level in one analysis and the .10 level in another.

This lack of consistency and significance may be in part a result of the small cell sizes resulting from sorting 41 or 46 subjects into six groups. There is additionally a lack of reliability of creativity groupings effected on the basis of scores on just one procedure.

Scores from the Instances and Uses procedures may be combined to provide a more stable "creativity" index, since these measures are substantially intercorrelated across the entire sample. A weighted sum of the four measures derived from these tests (uniqueness and total fluency for each of two procedures) was used, with a median split providing high and low groups. Two by two analyses of variance were performed, splitting IQ at the median for each sex. The corresponding two by two analysis was performed using the weighted sum derived from the Pictures test as the basis for creativity grouping, and forming IQ groups by a median split.

For females, there were no significant interactions in these analyses; and, with the exception of one _F_ significant at the .05 level, no main effects of creativity or intelligence. For males, however, a consistent pattern of results appeared for analyses using the Pictures test as the basis for grouping. Means are presented in Table XIX, and summaries of the analyses of variance in Table XX. For each of the three measures adminis-

Table XIX

Reflection--Impulsivity for Male Creativity--Intelligence Subgroups Based on Pictures Performance[a]

Measure	Group Hi Creative--Hi Intelligent (N=13)	Hi Creative--Lo Intelligent (N=7)	Lo Creative--Hi Intelligent (N=7)	Lo Creative--Lo Intelligent (N=14)
Dots	54.08	41.42	50.29	50.29
PPVT-A	54.85	42.00	48.86	50.00
PPVT-B	53.31	43.43	43.57	53.50
Matching Figures 2	54.85	41.57	53.86	47.71
Matching Figures 1	52.08	41.71	53.29	50.64

[a] Numbers in table are mean normalized latencies, with Mean=50 and S.D.=10 for all males.

Table XX

Reflection--Impulsivity for Male Subgroups Based on Pictures Performance

Summary of Analyses of Variance

Procedure	Source	df	MS	F	p
Dots	Intelligence	1	368.57	4.15	.05
	Creativity	1	59.12	0.67	
	Interaction	1	368.57	4.15	.05
	Within Cells	37	88.89		
PPVT-A	Intelligence	1	315.54	3.59	
	Creativity	1	9.32	0.11	
	Interaction	1	450.84	5.13	.05
	Within Cells	37	87.96		
PPVT-B	Intelligence	1	0.01	0.00	
	Creativity	1	0.26	0.03	
	Interaction	1	903.89	10.67	.01
	Within Cells	37	9.20		
Matching Figures 2	Intelligence	1	868.63	10.64	.01
	Creativity	1	61.19	0.75	
	Interaction	1	117.19	1.44	
	Within Cells	37	81.65		
Matching Figures 1	Intelligence	1	389.67	4.27	.05
	Creativity	1	236.75	2.60	
	Interaction	1	137.30	1.51	
	Within Cells	37	91.22		

tered under nonevaluative conditions, a significant interaction between creativity and intelligence is found. The interaction is attributable to highly impulsive performance by subjects in the group scoring high on creativity and low on intelligence. This group continues to respond impulsively on the Matching Figures procedures, but here the other group of subjects low on intelligence also tend to be impulsive, so that a significant effect of intelligence, rather than an interaction, is obtained.

A similar pattern of results is found in the analyses in which the Instances and Uses procedures are used as the basis for grouping. Here, however, none of the interactions are significant; and the main effect of intelligence is significant only for Matching Figures 2 ($\underline{F}$=10.92, df=1/37, $p < .01$). The means are presented in Table XXI, where in each of the five analyses the "high creative, low intelligence" subjects have latencies below the mean for the sample as a whole.

It had been predicted that subjects high on creativity and low on intelligence would show impulsive performance on measures obtained in an evaluation-laden context, but that they would be less impulsive under nonevaluative conditions. As in the correlational analyses reported in the preceding section, there is no evidence here that the administration context affects performance on reflection--impulsivity measures. No group showed a substantial or significant difference in mean latency of response between the two administration contexts.

The most straightforward prediction made for creativity--impulsivity relations was that subjects high on intelligence

Table XXI

Reflection--Impulsivity for Male Creativity--Intelligence Subgroups Based on Instances and Uses Performance[a]

	Group			
Measure	Hi Creative--Hi Intelligent (N=8)	Hi Creative--Lo Intelligent (N=12)	Lo Creative--Hi Intelligent (N=12)	Lo Creative--Lo Intelligent (N=9)
Dots	54.75	45.08	51.42	50.33
PPVT-A	50.25	48.17	54.42	46.22
PPVT-B	55.00	49.83	46.50	50.56
Matching Figures 2	58.13	46.08	52.08	45.33
Matching Figures 1	52.50	46.17	51.83	53.00

[a] Numbers in table are mean normalized latencies, with Mean=50 and S.D.=10 for all males.

and low on creativity would constitute the most reflective subgroup, and that they should be reflective under both administration contexts. Tables XIX and XXI provide no support for this hypothesis. In each table, males high on both creativity and intelligence measures are more reflective than those high only on intelligence, in four of five analyses (though these differences are small). For females, the same difference is found with all five measures for both creativity groupings; again, the differences are insubstantial.

The major relationships predicted between the creativity and impulsivity results, then, fail to be confirmed. The hypothesis of greatest reflectiveness for subjects high on intelligence but low on creativity was not supported by any significant Fs, and was contraindicated by such trends as there were in the data. The hypothesis of greatest impulsivity for subjects high on creativity and low on intelligence receives some support for males, though not for females. However, significant Fs appear only for creativity groups based on scores on the Pictures test; and then only when IQ grouping is accomplished by median split. Moreover, the predicted difference in performance for this group across contexts of test administration did not appear.

CHAPTER IV

DISCUSSION AND CONCLUSIONS

Reflection--Impulsivity

Clear positive results were obtained for the reflection--impulsivity dimension. Here, measures derived from Kagan's perceptual matching tasks intercorrelated strongly with several new measures which conformed to his definitional requirements but were phenotypically quite different from his tasks. Latencies and error scores exhibited negative correlations with one another, as in Kagan's studies, and showed correlations with intelligence comparable to his. Within the sample, age, sex, and IQ appeared not to affect these results. There is support, then, for Kagan's interpretation of reflection--impulsivity as a dimension of some importance and pervasiveness for childrens' performance; and there is evidence that it is a meaningful dimension even with relatively low intelligent children as young as five years.

There is no evidence that the context of test administration influences the individual's relative reflectiveness of impulsivity. Measures obtained under supposedly nonevaluative conditions correlated with those taken under evaluative pressure to the same degree as they correlated with one another. These correla-

tions were somewhat lower than those between the two measures obtained under evaluative conditions, but this difference is likely the result of greater reliability for the latter tasks. It remains possible, however, that intercorrelations of measures will be higher in general when the evaluational features of the situation are made salient. There could be, then, a difference between contexts in which the contribution of reflection--impulsivity to task performance is greater when adequacy of response is at issue than when it is not. Latencies would be more variable but not systematically different in a permissive atmosphere. This possibility cannot be tested on the present data, since no test was given under both administration contexts.

It will be recalled that four possible bases for individual differences in reflection--impulsivity have been proposed (pp. 60-62). This study provides a little evidence relevant for the choice among them. First, there is the possibility that greater reflectiveness is associated with greater cognitive maturity, and thus represents level of development rather than style of information processing. On this hypothesis, one can predict positive correlations between reflection and both age and intelligence. Neither relation was found--age differences in reflectiveness did not appear, and the correlations between latencies and IQ were generally nonsignificant. However, these failures of relation are not conclusive. It may be argued that the age range in the present sample was too small to show the real differences that exist. The IQ range was large; but here it is possible that verbal intelligence is not the appropriate measure

of cognitive development--measures like those employed by Piaget, or even a performance IQ as assessed by standard intelligence tests, might differentiate impulsive from reflective individuals.

Two additional proposals are based on the consideration of specific motivational variables. The impulsive child is seen as anxious over the adequacy of his performance, responding too quickly in order to terminate anxiety; or the reflective child is seen as highly motivated to achieve, withholding response to avoid making errors. On either of these interpretations, it may be predicted that the evaluative context of test administration will influence performance. When evaluation of response adequacy is made salient, individuals threatened by evaluation should become more impulsive, or individuals desiring to achieve should become more reflective, relative to their performance under nonevaluative conditions.

No direct test of these predictions is possible, since no single measure of reflection was obtained under both the permissive and the evaluative conditions, and response times are not comparable across tests. The intercorrelations of the measures provide no definitive information, since the pattern of relations is compatible with either a lack of systematic change in individual performance with administration context, or a systematic change in which either or both extremes of the distribution become more extreme--either of these conditions could lead to no change in magnitude of correlations when those across contexts are compared with those within the permissive context. The high correlations between the two tests administered

in an evaluative context, as discussed earlier, may be simply a function of greater test reliability (either because of test content or because of the larger number of items on which the total score is based); or it could arise because of a larger performance difference between "really reflective" and "really impulsive" individuals.

It is possible, however, to consider performance differences within the two measures obtained under evaluative conditions. Assuming that the evaluational context is most salient to the child on trials immediately following an error (since the experimenter informs him he has made an error and will receive no marble for that trial), it can be predicted that reflective individuals will be more reflective following errors than following correct choices, if reflection represents the desire to achieve; while impulsive individuals will be more impulsive following errors than following correct choices, if impulsivity represents anxiety over adequacy of performance. For these analyses, reflective and impulsive individuals were chosen by median split of mean latency on the three measures obtained in the nonevaluative testing context. Data for males were combined with those for females, after the median split was made, to yield 41 reflective and 44 impulsive individuals. Unlike all previous analyses, raw rather than normalized latencies were employed.

In each test, the null hypothesis was that the proportion of individuals showing longer latencies on trials following erroneous choices than on trials following correct choices would

not differ from .50. For individuals identified as reflective, the null hypothesis could not be rejected: On MF2, 59% of the subjects had longer latencies following errors ($z=1.11$, n.s.); for MF1, 63% of the changes were in this direction ($z=1.78$, n.s.). This test also is inconclusive: it could be that the major part of the variance is carried by persons rather than situations; or that evaluational cues are highly salient for reflective individuals throughout these procedures, leading to a "ceiling" effect.

Turning to those individuals classified as impulsive, there is at last some evidence which is unambiguous. If anxiety over adequacy of performance is important for them, they should be most impulsive when evaluational cues are most salient. A significant proportion of these individuals, however, showed longer latencies on trials following errors than on trials following correct responses. On MF2, 86% of the impulsives became less impulsive ($z=7.03$, $p < .0001$), and on MF1, 74% changed toward less impulsivity ($z=3.51$, $p < .0005$), after errors. Thus, the anxiety hypothesis is untenable as an explanation of differences between reflective and impulsive individuals.

The fourth proposal made by Kagan for a basis for individual differences on this dimension was that cognitive impulsivity is a special case of a more general disposition, reflected for example in high gross motor activity and high physical risk taking. Some indication of generality for the dimension is provided in this study--individual differences are consistent across a range of tasks greater than that of the perceptual-matching

procedures employed by Kagan; and these differences appear before the child has been subjected to the achievement and anxiety stresses of the classroom. The proposal of a constitutional base for this generality has of course not been touched on by the evidence available here.

It is clear that direct study, rather than indirect argument, is needed. Kagan's support for the interpretations he has offered is as inconclusive as that presented here. Anxiety over adequacy of performance appears to be inadmissible as an explanation of the differences found; none of the remaining possibilities deserves either acceptance or rejection on the basis of present information. The best guess that can be made, however, is that the dimension has a degree of generality not derivable from the specific motivational or maturational variables that have been considered.

Interpretations of Creativity Results

If the Pictures test could be ignored, there would be a strong basis for claiming that the creativity dimension measured by Wallach and Kogan was reproduced with five and six year old children, despite the differences in age, intelligence, and social class between their sample and the present one. The characteristics and correlates of performance on the Instances and Uses procedures appear to be completely consistent with this conclusion. Unique and nonunique response totals, for example, were highly correlated within and between procedures.

They had low and even negative correlations with the intelligence measure, although use of a mental age index on a test of verbal comprehension gave maximum opportunity for positive correlations between "creativity" and "intelligence." Within the procedures, high fluency totals were associated with a greater percent of unique responses, and were not derived by exhaustive use of highly common categories. Results were similar across sex, and showed no great differences for younger versus older or for brighter as compared to duller children.

Ignoring the Pictures data, it would also be concluded that the predictions regarding relationships of reflection--impulsivity to creativity and intelligence were in error. These data provided none of the significant interactions that had been predicted: for males, there was one weak trend which was partially consistent with expectations (for highly creative, low intelligent children); but this trend showed no significance and not even complete self-consistency.

Both positive and negative conclusions, then, would be clear, though not necessarily correct. Data from the Pictures test, however, force consideration of several alternatives, with reliance on evidence which is more indirect than is to be desired. Because of this indirectness, and the small numbers of subjects involved when considering subgroups, interpretations can only be regarded as suggestive. Once this reservation is made, it will be most useful for future research efforts to consider freely the possibilities raised by the data and the evidence that is relevant for each.

The first possibility is that the Instances and Uses measures are tapping the creativity dimension studied by Wallach and Kogan, and that the failure of the Pictures measures to correlate with these follows from the inappropriateness of the test rather than from lack of a unitary dimension. It is clear, as indicated above, that Instances and Uses are measuring something in common. Many characteristics of the data from these procedures are consistent with those anticipated if what they are measuring is the creativity dimension. Their correlations with time in task, taken as a measure of task involvement, appear also to be compatible with this interpretation.

Moreover, a number of plausible explanations can be given for the failure of Pictures measures to correlate with those from Instances and Uses, despite the presence of a unitary creativity dimension. One possibility which was raised earlier was that high response totals on Pictures demand greater convergent and divergent thinking ability than do high totals on the other procedures. If this were so, Pictures might correlate with the other procedures only with a high ability level, so that performance is not mainly a function of abstractive ability. Support was found for males, where substantial correlations with both MA and the other divergent thinking measures were found when low intelligence subjects were eliminated. For females, however, Pictures performance was not strongly related to any other performance either across all subjects or for the more intelligent subjects alone.

This failure for females might result from social-

psychological rather than cognitive-structural variables. It has been often remarked and occasionally demonstrated (e.g., McClelland et al, 1953; Sears, P., 1962) that males show relatively greater task involvement; females, greater orientation toward approval or affiliation. If such a difference were initially present in these children, it might have been manifested in performance on Pictures, the first creativity procedure given, while becoming unimportant on the second and third creativity procedures because of increasing familiarity with the game and the examiner. In short, one motivational syndrome might initially have been relevant for females, while a different syndrome was of importance for performance of females in the two remaining tasks and of males in all three tasks. No evidence, however, supports this possibility. There are no sex differences in response totals for Pictures, and there are correlations on the order of .5 between time measures, interpreted as measures of task involvement, from the various creativity tasks.

It has also been proposed that performance for females is more a function of the form of the test (visual or verbal stimuli, written or spoken responses, etc.), while for males it is more a function of kind of information processing required (Guilford, Merrifield, and Cox, 1961). Here the major difference in task form is that between Pictures and the other procedures, with abstract pictorial stimuli in the former and verbal categories or concrete objects presented in the latter. Additionally, however, responses to Pictures and Instances may share the feature that the appropriate response is a single word

which names an instance of a category; while for Uses responses are more often phrases and, if "uses for a newspaper" is considered a category, it is a functional one rather than one involving superordinate-subordinate relationships. If this latter difference affects the degree of relationship between tasks for females and not for males, some plausibility will be given to the proposal that the figural-verbal or abstract-concrete difference between Pictures and the other procedures might also have done so.

For all females, mean correlation among the unique and non-unique totals for Pictures and the corresponding totals for Instances is .29 ($p \approx .05$); with Uses it is .05. The corresponding mean correlations across all males are .08 with Instances and .05 with Uses. The difference for females is consistent with the interpretation, though not significant; while there is no difference for males.

There is some plausibility, then, for the proposal that task form may have influenced performance on Pictures for females while not for males. This explanation may be coupled with the notion that for males Pictures performance is a joint function of creativity and intelligence with the latter predominating in low IQ children. Together, these interpretations would explain the failures of correlation between Pictures and the other creativity procedures, despite the presence and unity of the creativity dimension studied by Wallach and Kogan.

An alternative interpretation of the creativity results is that a unitary creativity dimension is not present, but rather, "visual" and "verbal" procedures measure independent abilities

within the divergent thinking domain. Guilford has suggested that tests with figural content will load on factors parallel to, but distinct from, those loaded with verbal (semantic) content (Guilford, 1959a). He has not used such tests, however. Torrance has used both verbal and nonverbal procedures, reporting low correlations between domains; but in his studies, unlike the present one, tasks with nonverbal stimuli have nonverbal responses as well. (Torrance, 1963) Wallach and Kogan provide the only fully comparable data; in their study, no visual-verbal separation appeared.

There may well be differences with both age and intelligence in the figural-verbal relation. It should be noted, however, that the present difference does not take the direction predicted by the hypothesis of increasing mental "differentiation" with age. By that hypothesis, cognitive dimensionality should increase with age (and with intelligence). Here, it appears that a single process serves for older children in place of what are distinct and relatively independent processes in younger children.

It is tempting to speculate on how a course toward decreasing dimensionality could proceed. One possibility is that various pressures during development lead to the imposition of greater constraint (e.g., motivational and stylistic tendencies) on intellective functioning, so that such differences as appear in the younger children are later of relatively minor importance in the total matrix of determinants of performance. Another is that storage or retrieval processes develop so that, for example, the relevant associative hierarchies for various kinds of tasks

have greater similarity in the older than in the younger child. A third is that figural and verbal tasks require different subprocesses to somewhat different degrees, and that, so long as these subprocesses are immature, performance on different tasks will differ. When a subprocess has reached a level of efficiency sufficient for the tasks contrasted, however, it no longer plays a differentiating role in performance. This latter is another proposal of a crucial threshold level, analogous to the hypothesis that creativity and intelligence are independent above some minimal level of intelligence; here, however, development with age is seen as important.

These possibilities are researchable, though not easily operationalized. For the present it can only be pointed out that a "real" figural-verbal distinction is possible in the children studied, and that such a distinction in five and six year old children is not incompatible with the absence of this distinction in older or brighter children.

Two interpretations of the creativity results have been offered. The first proposes that the Instances and Uses procedures measure the creativity dimension studied by Wallach and Kogan, and that the Pictures test fails to correlate with these because of specific characteristics of the test. The second suggests that a "real" figural-verbal distinction is present in five and six year old children, and therefore no unitary dimension comparable to that seen in older children is to be found. Some plausibility may be developed for a third alternative; namely, that it is only the Pictures procedure which taps the dimension

studied in older children, while Instances and Uses are related to a more primitive aspect of the child's functioning.

Argument for this position might start by noting that both Instances and Uses provide the subject an opportunity to achieve high response totals by use of many associates which do not represent new or unique associations *for him*. With familiar categories on Instances, and common objects on Uses, the subject may recount present perception (for Instances) and past experience (for both procedures). His scores will then depend in part on the richness of his experience, and in part on variables determining how much of this experience he is willing or able to recount. Recounting will be influenced by the thoroughness of his search and by the efficiency of storage. Most of these variables are not ones which are thought of in terms of creativity.

With the older children studied by Wallach and Kogan, it may be that responses do represent recombinations of experience, and that unique responses for the sample are likely to be genuinely novel ones for the subject--that is, not drawn directly from his experience, but constructed from it. Older children may accept as an implicit part of the task that new and genuinely different responses should be given; and that, for example, it is not fair play to name five different kinds of balls for round things. Such self-imposed censoring seems to be weak or absent in the younger children, where many responses fall into highly conventional categories, and many are given which appear to be bizarre.

Only on the Pictures test is it impossible for high scores

to represent how much is drawn unaltered from memory. Categories here are abstract and unfamiliar, so that it is unlikely that any list of exemplars is organized in memory or can be recovered by scanning limited portions of storage. It is quite likely, then, that a different form of search through memory is required than that for Instances and Uses. The fit of response to stimulus, also, may be different--the exemplar must conform to essential properties of the stimulus; it can be equivalent to the stimulus, but not identical to it, since nonessential properties must be disregarded. The child, of course, must determine what are the essential properties of stimulus and potential responses.

The self-imposed censoring engaged in by older children may render the tasks all similar, in that these children do not take the opportunity provided by some procedures for extensive unaltered listing from memory or perception. Therefore, they may respond to all tasks by use of the same processes as are required by Pictures. For younger children, on the other hand, there are indications that censoring is weak and different processes may be operating.

It follows from these arguments that Instances and Uses may measure a dimension different from, though related to, that tapped in older children. Their correlations with time in task, and the intercorrelations of time measures, suggest that this dimension may involve a disposition toward creativity. With tasks requiring no production of a "new" response, but only a search for items meeting some criterion, this disposition is manifested in high response totals as well as in long involvement

in the task; but when genuinely new integrations are required, it is sufficient only to produce involvement, not to produce high response totals.

There is no direct evidence that what is measured in the younger children is basic for, or related to, the variable studied with older ones. Moreover, no evidence has been cited which compels the conclusion that the Pictures test taps what is measured in older children. The only basis for such a conclusion lies in the significant interactions for reflection--impulsivity when IQ and Pictures performance are the bases for grouping. Despite the high correlation between Pictures scores and intelligence, the small group of males high on intelligence and low on Pictures totals formed the group of males most impulsive on all five measures. This finding was not completely consistent with expectations but is compatible with Wallach and Kogan's descriptions. With the Instances and Uses procedures, the corresponding analysis produced no significant interactions, merely a weak trend corresponding in direction to that for Pictures. The weakness of the finding, especially considering its failure to appear in females, cannot be overemphasized; but it is most easily explained as showing the beginning of appearance of the dimension found in older children.

It has been suggested that some data are consistent with each of three possibilities: a unitary creativity dimension is not to be found in children of kindergarten age; the dimension is present and is measured by the Instances and Uses procedures, but not by the Pictures test; the dimension is just beginning

to appear in the males of the sample, as seen in the Pictures test, while Instances and Uses are tapping a disposition which may be more basic. Longitudinal study will be important in discriminating among these possibilities. Just as important will be the examination of relationships with many of the variables employed by Wallach and Kogan.

Conclusions

The present experiment represents a "focus gambling" approach to problem solving, since both characteristics of the subjects and variables to be related to the creativity measures were varied from the preceding "positive instance," the study by Wallach and Kogan. It will be necessary to examine the effects of age and intelligence alone before definitive conclusions as to the role of these variables can be reached. Likewise, replication with older subjects will be necessary before the fate of the "cognitive control" interpretation is entirely decided. However, some conclusions can be drawn with greater certainty than may be indicated by the number of alternative interpretations outstanding.

First, it is clear that, however the "creativity" measures are related to the dimension studied by Wallach and Kogan and to creativity in the sense of significant production, they are not simply unusual or low-quality measures of intelligence. Instances and Uses had no significant positive correlations with a verbal comprehension measure, either across a range from below

average to well above average intelligence, or for a subsample of brighter individuals. Pictures showed positive correlations with mental age only for males, and showed additional correlations not derivable from its relation to intelligence.

There is strong evidence in these data that verbal tests of ideational fluency share a major part of their variance with one another. Many correlations are .5 or higher, whether the measure used is total responses, total relevant responses, or a breakdown into unique and nonunique responses; and for each sex, overall or separately for subjects of higher or lower intelligence. It is clear, then, that these tests measure something in common, something other than intelligence. Ambiguity resides in interpretations of their failure to relate to another ideational fluency measure, which differed from them in several possibly important features: in requiring greater abstractive ability; in being less susceptible to response production through mere recall; in using figural rather than verbal stimuli (or concrete stimuli which could easily have been presented verbally); and in being the first of the creativity procedures administered. The alternative interpretations of the creativity results largely reflect assignment of different weights to these differences.

There is justification for the position (Thorndike, 1963) that, in the absence of evidence that the fluency procedures are tapping the creativity dimension studied by Wallach and Kogan, a less pretentious term than "creativity" should be employed. Labeling them as measures of "verbal ideational fluency" and "figural ideational fluency" would involve less inference as to

the role of what they measure in the cognitive processes; but it would not aid in discovering this role. It seems extremely unlikely that what they measure bears no relation to the creativity dimension found in older children. This conclusion is based on the consistencies of characteristics of data from these procedures with expectations based on Wallach and Kogan's findings; and is supported by relationships with time spent in task, interpreted as a measure of involvement specific to this kind of task. There is no certainty, however, as to what is the form of their relation to the creativity dimension. The alternatives that have been considered are of two classes. First, they may measure the same dimension as is found in older children. "The same dimension" here is most simply interpreted to mean that they have the same pattern of correlates with other variables that can be measured at each age. Second, they may measure a dimension which is basic to the dimension found later. "Basic to" can mean either "prior to," in that the development of some processes which change with age will be sufficient to produce the correlational patterns found by Wallach and Kogan; or it may mean "necessary but not sufficient," if for example a high level of intelligence as well as development with age is necessary for the appearance of these correlates.

It may be noted that little attention has been given to the possibility that intelligence level, rather than development correlated with age, was crucial in determining the pattern of correlational results. This possibility should not be ignored. However, it is unlikely that a higher mean IQ for the sample

would have been sufficient to produce grossly different results. In each sex, mean IQ for groups from which low intelligence subjects were deleted was about 111. This is lower than the cutoff level of 120 which has been proposed by MacKinnon, Roe, and others; but it is sufficiently high that many subjects have IQs close to or above 120. Note, also, that the cutoff level proposed is one below which creativity and intelligence measures will be substantially correlated, while above this level they should show independence. It is clearly not the case here that divergent and convergent thinking are correlated for low intelligence children and that the correlation decreases with increasing IQ. Rather, differences are in the direction that creativity measures hold less variance in common for subjects of low intelligence, and that their intercorrelations increase with increasing IQ.

As in Wallach and Kogan's study, no direct test was made of the assumption that the use of a permissive, game-like context for creativity measurement is crucial for results. However, research reviewed in the introductory section of this paper finds positive correlations between convergent and divergent thinking measures, for children as well as for adults, in those studies in which no explicit effort to produce this atmosphere was made. The administration context provides the most obvious distinction between studies finding positive correlations with IQ and studies not finding such correlations; there is fair support for the assumption, then, but direct study of the role of context is called for.

No conclusive evaluation of the "cognitive control" model

can be made. Small sample size, especially considering the possibility of sex differences, would have rendered this evaluation only tentative at best. Given the ambiguity in the creativity results, there is little justification for even tentative conclusions. Such evidence as there is, however, offers no support for the model. Its strongest prediction was that individuals high on intelligence and low on creativity should have the longest latencies on measures of reflection--impulsivity. This did not appear when the Pictures procedure formed the basis for creativity grouping, or when the Instances and Uses procedures were used for this purpose. Furthermore, groups formed by splitting subjects into those spending longer and shorter time in responding on the creativity procedures failed to show greater reflection for these subjects. In all of these analyses, rather, the most reflective subjects were those high on both creativity and intelligence measures. This finding was unpredicted and was not supported by statistical significance, but is at least suggestive for future research.

The model also predicted greatest impulsivity for subjects high on creativity but low on intelligence, and that this impulsivity would appear only for measures obtained when cues for evaluation of adequacy were salient. There was no evidence that the evaluative as opposed to permissive context influenced impulsivity measures; and only one relatively unstable finding of greater impulsivity for this group of subjects. If the Instances and Uses procedures are measuring the "same" dimension as that found in older children, then, the hypothesis is disconfirmed.

If either of the other explanations of the creativity results is correct, however, the assumption has not been adequately tested. Results involving the Pictures test suggest that, for older subjects for whom Pictures performance is related to other creativity scores and not to IQ, the group high on creativity and low on intelligence will show impulsivity.

Finally, it may be reiterated that there was no difficulty in replicating Kagan's reflection--impulsivity dimension, or in showing its relevance to a wider range of decision tasks than those in which he has measured it. Since these children are tested before they have been subjected to the anxieties and achievement stresses of the school environment, this success increases the plausibility of one of the four alternative explanations for individual differences in reflection--impulsivity that Kagan has suggested--namely, that performance in these cognitive judgmental tasks is one facet of a broader disposition manifested, for example, in level of gross motor activity and physical risk taking. How pervasive this disposition may be is not yet clear, but at least it can be generalized beyond perceptual matching tasks.

REFERENCES

REFERENCES

Ach, N. Ueber die willenstaetigkeit und das denken. Goettingen: Vandanhoeck and Ruprecht, 1905. See D. Rapaport (Ed.), Organization and pathology of thought. New York: Columbia Univer. Press, 1951. Pp. 15-39.

Asher, J.J. Towards a neo-field theory of problem solving. J. gen. Psychol., 1963, 68, 3-8.

Barron, F. The disposition towards originality. In C. Taylor (Ed.), The 1955 University of Utah conference on the identification of creative scientific talent. Salt Lake City: Univer. Utah Press, 1956. Pp. 156-170.

Bousfield, W.A., & Barclay, W.A. The relationship between order and frequency of occurrence of restricted associative responses. J. exp. Psychol., 1950, 40, 643-647.

Bowers, J. Explorations of creative thinking in the early school years: XIV. A preliminary factor-analytic study of the creative thinking abilities of children. Minneapolis: Univer. Minn., Bureau of Educ. Res., 1960.

Braine, M.D.S. The ontogeny of certain logical operations: Piaget's formulation examined by nonverbal methods. Psychol. Monogr., 1959, 73, No. 5, Whole No. 475.

Brown, R.W., & Berko, Jean. Word association and the acquisition of grammar. Ch. Develop., 1960, 31, 1-14.

Carroll, J.B. A factor analysis of verbal abilities. Psychometrika, 1941, 6, 297-307.

Cattell, R.B. The description and measurement of personality. New York: World Book, 1946.

Cattell, R.B. The culture free intelligence test. Champaign, Ill.: Institute for Personality and Ability Testing, 1950.

Christensen, P.R., Guilford, J.P., & Wilson, R.C. Relations of creative responses to working time and instructions. J. exp. Psychol., 1957, 53, 82-88.

Cline, V.B., Richards, J.M., Jr., & Abe, C. The validity of a battery of creativity tests in a high school sample. *Educ. psychol. Measmt*, 1962, *22*, 781-784.

Cline, V.B., Richards, J.M., Jr., & Needham, W.E. Creativity tests and achievement in high school science. *J. appl. Psychol.*, 1963, *47*, 184-189.

Cooperative Test Division. *Cooperative school and college ability tests: Technical report.* Princeton: Educational Testing Service, 1957. (a)

Cooperative Test Division. *SCAT: Directions for administering and scoring.* Princeton: Educational Testing Service, 1957. (b)

Cooperative Test Division. *Cooperative sequential tests of educational progress: Technical report.* Princeton: Educational Testing Service, 1957. (c)

Cooperative Test Division. *STEP: Directions for administering and scoring.* Princeton: Educational Testing Service, 1957 (d)

Cooperative Test Division. *Cooperative sequential tests of educational progress: Teacher's guide.* Princeton: Educational Testing Service, 1959.

Crutchfield, R. The creative process. In *Conference on the creative person.* Berkeley: University of California, Institute of Personality Assessment and Research, 1961. Ch. 6.

Dunn, L.M. *The Peabody picture vocabulary test.* Nashville, Tenn.: American Guidance Service, Inc., 1959.

Flavell, J.H. *The developmental psychology of Jean Piaget.* Princeton: Van Nostrand, 1963.

Flavell, J.H., Draguns, J., Feinberg, L.D., & Budin, W. A microgenetic approach to word association. *J. abnorm. soc. Psychol.*, 1958, *57*, 1-7.

Flescher, I. Anxiety and achievement of intellectually gifted and creatively gifted children. *J. Psychol.*, 1963, *56*, 251-268.

Freud, S. *Leonardo da Vinci: A study in psychosexuality.* New York: Random House, 1947.

Freud, S. The relation of the poet to day-dreaming. In *Collected papers of Sigmund Freud.* Vol. 4. New York: Basic Books, 1959. Pp. 173-183. (a)

Freud, S. Formulations regarding the two principles in mental functioning. In *Collected papers of Sigmund Freud*. Vol. 4. New York: Basic Books, 1959. Pp. 13-21. (b)

Gamble, A.O. Suggestions for future research. In C. Taylor (Ed.), *The 1959 University of Utah research conference on the identification of creative scientific talent*. Salt Lake City: Univer. Utah Press, 1959. Pp. 292-297.

Getzels, J.W., & Jackson, P.W. *Creativity and intelligence: Explorations with gifted children*. New York: Wiley, 1962.

Gewirtz, J.L. Studies in word fluency. I. Its relation to vocabulary and mental age in young children. *J. genet. Psychol.*, 1948, *72*, 165-176. (a)

Gewirtz, J.L. Studies in word fluency. II. Its relation to eleven items of child behavior. *J. genet. Psychol.*, 1948, *72*, 177-184. (b)

Ghiselin, B. *The creative process*. Berkeley: Univer. Calif. Press, 1952.

Golann, S.E. Psychological study of creativity. *Psychol. Bull.*, 1963, *60*, 548-565.

Goldstein, K. *The organism*. New York: American Book, 1939.

Guilford, J.P. The structure of intellect. *Psychol. Bull.*, 1956, *53*, 267-293.

Guilford, J.P. Creative abilities in the arts. *Psychol. Rev.*, 1957, *64*, 110-118.

Guilford, J.P. Three faces of intellect. *Amer. Psychol.*, 1959, *14*, 469-479. (a)

Guilford, J.P. Traits of creativity. In H. Anderson (Ed.), *Creativity and its cultivation*. New York: Harper, 1959. Pp. 142-161. (b)

Guilford, J.P., Frick, J.W., Christensen, P.R., & Merrifield, P.F. A factor-analytic study of flexibility in thinking. *Univer. Sth. Calif. Psychol. Lab. Rep.*, 1957, No. 18.

Guilford, J.P., Kettner, N.W., & Christensen, P.R. A factor-analytic study across the domains of reasoning, creativity, and evaluation. II. Administration of tests and analysis of results. *Univer. Sth. Calif. Psychol. Lab. Rep.*, 1956, No. 16

Guilford, J.P., Merrifield, P.R., & Cox, Anna B. Creative thinking in children at the junior high school levels. *Univer. Sth. Calif. Psychol. Lab. Rep.*, 1961, No. 26.

Hadamard, J. The psychology of invention in the mathematical field. New York: Dover, 1954.

Harmon, L.R. The development of a criterion of scientific competence. In C. Taylor (Ed.), The 1957 University of Utah conference on the identification of creative scientific talent. Salt Lake City: Univer. Utah Press, 1958. Pp. 82-97.

Harvey, O.J., Hunt, D.E., & Schroder, H.M. Conceptual systems and personality organization. New York: Wiley, 1961.

Hollingshead, A.B. Two factor index of social position. A.B. Hollingshead, 1965. Yale Station, New Haven. Copyright 1957.

Holt, R.R., & Havel, Joan. A method for assessing primary and secondary process in the Rorschach. In Maria A. Rickers-Ovsiankina (Ed.), Rorschach Psychology. New York: Wiley, 1960. Pp. 263-315.

Iscoe, I., & Pierce-Jones, J. Divergent thinking, age, and intelligence in white and Negro children. Ch. Develop., 1964, 35, 785-798.

Jackson, P., & Messick, S. The person, the product, and the response: conceptual problems in the assessment of creativity. Mimeo., 1964.

Johnson, D.M., & Reynolds, F. A factor analysis of verbal ability. Psychol. Rec., 1941, 4, 183-195.

Kagan, J. Developmental studies of reflection and analysis. Mimeo., 1964.

Kagan, J., & Moss, H.A. Birth to maturity. New York: Wiley, 1962.

Kagan, J., Rosman, Bernice L., Day, Deborah, Albert, J., & Phillips, W. Information processing in the child: Significance of analytic and reflective attitudes. Psychol. Monogr., 1964, 78, No. 1, Whole No. 578.

Kendler, T.S., Kendler, H.H. & Wells, D. Reversal and nonreversal shifts in nursery school children. J. Comp. physiol. Psychol., 1960, 53, 83-88.

Kris, E. Psychoanalytic explorations in art. New York: International Universities Press, 1952.

Kubie, L.S. Neurotic distortion of the creative process. Lawrence: Univer. of Kansas Press, 1958.

Luria, A.R., & Yudovich, F. Ia. Speech and the development of mental processes in the child. London: Staples Press, 1959.

MacKinnon, D.W. The highly effective individual. Teachers' Coll. Rec., 1960, 61, 367-378.

MacKinnon, D.W. The nature and nurture of creative talent. Amer. Psychol., 1962, 17, 484-495.

McClelland, D., Atkinson, J., Clark, R., and Lowell, E. The achievement motive. New York: Appleton-Century-Crofts, 1953.

McNemar, Q. Lost: our intelligence? Why? Amer. Psychol., 1964, 19, 871-882.

Mednick, S.A. The associative basis of the creative process. Psychol. Rev., 1962, 69, 220-232.

Miller, G.A., Galanter, E., & Pribram, K.H. Plans and the structure of behavior. New York: Holt, 1960.

Mooney, R.L. A conceptual model for integrating four approaches to the identification of creative talent. In C.W. Taylor & F. Barron (Eds.), Scientific creativity: its recognition and development. New York: Wiley, 1963. Pp. 331-340.

Myden, W. Interpretation and evaluation of certain personality characteristics involved in creative production. Percept. mot. Skills, 1959, 9, 139-158.

Newell, A., Shaw, J.C., & Simon, H.A. Elements of a theory of human problem solving. Psychol. Rev., 1958, 65, 151-166.

Newell, A., Shaw, J.C., & Simon, H.A. The processes of creative thinking. In H.E. Gruber, G. Terrell, & M. Wertheimer (Eds.), Contemporary approaches to creative thinking. New York: Atherton Press, 1962. Pp. 63-119.

Piaget, J. The language and thought of the child. New York: Harcourt, Brace, 1926.

Piaget, J. Judgment and reasoning in the child. New York: Harcourt, Brace, 1928.

Piers, Ellen V., Daniels, Jaqueline M., & Quackenbush, J.F. The identification of creativity in adolescents. J. educ. Psychol., 1960, 51, 346-351.

Pine, F., & Holt, R.R. Creativity and primary process: a study of adaptive regression. J. abnorm. soc. Psychol., 1960, 61, 370-379.

Reese, H.W. Verbal mediation as a function of age level. Psychol. Bull., 1962, 59, 502-509.

Roe, Anne. Crucial life experiences in the development of scientists. In E.P. Torrance (Ed.), Education and talent. Minneapolis: Univer. Minn. Press, 1960. Pp. 66-77.

Rogers, C.R. Toward a theory of creativity. Etc., 1954, 11, 249-260.

Schactel, E.G. Metamorphosis. New York: Basic Books, 1959.

Schaefer, E.A., & Bayley, Nancy. Maternal behavior, child behavior, and their intercorrelations from infancy through adolescence. Monogr. Soc. Res. Child Develop., 1963, 28, Serial No. 87, Whole No. 3.

Sears, Pauline S. Correlates of need Achievement and need Affiliation and classroom management, self-concept, achievement and creativity. Unpublished manuscript, Laboratory of Human Development, Stanford University, 1962.

Sherman, Vivian S. Personality correlates of differential performance on intelligence and creativity tests. Unpublished Ph.D. dissertation, Stanford University, 1964.

Snedecor, G.W. Statistical methods (4th ed.). Ames, Iowa: Iowa State Coll. Press, 1946.

Taylor, C.W. A factorial study of fluency in writing. Psychometrika, 1947, 12, 239-262.

Taylor, C.W. (Ed.). The 1959 University of Utah research conference on the identification of creative scientific talent. Salt Lake City: Univer. Utah Press, 1959.

Taylor, D.W. Variables related to creativity and productivity among men in two research laboratories. In C.W. Taylor & F. Barron (Eds.), Scientific creativity: its recognition and development. New York: Wiley, 1963. Pp. 228-250.

Taylor, I.A. The nature of the creative process. In P. Smith (Ed.), Creativity. New York: Hastings House, 1959. Pp. 51-82.

Terman, L.M., & Merrill, Maud A. Measuring intelligence. Boston: Houghton Mifflin, 1937.

Thorndike, R.L. Some methodological issues in the study of creativity. In Proceedings of the 1962 invitational conference on testing problems. Princeton: Educational Testing Service, 1963. Pp. 40-54.

Thurstone, L.L. A factorial study of perception. Psychometric Monogr., No. 4. Chicago: Univer. Chicago Press, 1944.

Thurstone, L.L., & Thurstone, Thelma G. Factorial studies of intelligence. Psychometric Monogr., No. 2. Chicago: Univer. Chicago Press, 1941.

Torrance, E.P. Educational achievement of the highly intelligent and the highly creative: Eight partial replications of the Getzels-Jackson study. Minneapolis: Univer. Minn., Bureau of Educ. Res., 1960.

Torrance, E.P. Explorations in creative thinking in the early school years: a progress report. In C.W. Taylor and F. Barron (Eds.), Scientific creativity: its recognition and development. New York: Wiley, 1963. Pp. 173-183.

Vygotsky, L.S. Thought and language. New York: The M.I.T. Press and Wiley, 1962.

Wallach, M.A. Research on children's thinking. In H.W. Stevenson (Ed.), Child psychology: The sixty-second yearbook of the National Society for the Study of Education. Chicago: Univer. Chicago Press, 1963. Pp. 236-276.

Wallach, M.A., & Kogan, N. A new look at the creativity-intelligence distinction. Mimeo., 1964.

Wallach, M.A., & Kogan, N. Modes of thinking in young children: A critical examination of the creativity-intelligence distinction. New York: Holt, Rinehart and Winston, 1965.

Wallas, G. The art of thought. New York: Harcourt, Brace, 1926.

Wechsler, D. Wechsler intelligence scale for children: Manual. New York: Psychological Corp., 1949.

White, S.H. Evidence for a hierarchical arrangement of learning processes. In L.P. Lipsitt & C.C. Spiker (Eds.), Advances in child development and behavior. Vol. 2. New York: Academic Press, 1965. Pp. 187-220.

Wilson, R.C., Guilford, J.P., Christensen, P.R., & Lewis, D.J. A factor-analytic study of creative thinking abilities. Psychometrika, 1954, 19, 297-312.

Witkin, H.A., Dyk, R.B., Faterson, H.F., Goodenough, D.R., & Karp, S.A. Psychological differentiation. New York: Wiley, 1962.

Woodrow, H., & Lowell, Frances. Children's association frequency tables. Psychol. Monogr., 1916, 22, No. 5, Whole No. 97.

Woodworth, R.S., & Schlosberg, H. Experimental psychology. (2nd ed.) New York: Holt, 1954.

Yamamoto, Kaoru. Educational achievement of highly creative students: threshold of intelligence. Minneapolis: Univer. Minn., Bureau of Educ. Res., 1961.

Yamamoto, Kaoru. Further analysis of the role of the creative thinking abilities in high school achievement. Minneapolis: Univer. Minn., Bureau of Educ. Res., 1962.

BIOGRAPHY

NAME: William Cornelius Ward, Jr.

BORN: Baltimore, Maryland; November 21, 1939

DEGREE EARNED: A.B., The Johns Hopkins University, June, 1961

ACADEMIC HONORS: Phi Beta Kappa

FELLOWSHIPS HELD:

1961-63	James B. Duke Fellowship
1963-64	National Science Foundation Cooperative Graduate Fellowship
1964-65	National Science Foundation Open Graduate Fellowship

POSITIONS AND APPOINTMENTS:

1962, 1963 (Summer)	Senior Technical Assistant, Bell Telephone Laboratories
1965 (Summer)	Associate Professor of Education, North Carolina College
1965-Present	Assistant Professor of Psychology, Stanford University

PUBLICATIONS:

Jenkins, H.M., & Ward, W.C. The judgment of contingency between responses and outcomes. Psychol. Monogr., 1965, 79, No. 1 (Whole No. 594).

Ward, W.C. Concepts au Geneva. Review of J.G. Wallace, Concept growth and the education of the child. Contemp. Psychol., in press.

Ward, W.C., & Jenkins, H.M. The display of information and the judgment of contingency. Canad. J. Psychol., 1965, 19, 231-241.

D. 1966
In

6/15/67